

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:31:50 ; Search time 246 Seconds  
(without alignments)

8651.818 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 atgtttacacgtcgggcac.....ccgcgctcagggtcgctgc 501

Scoring table: OLIGO NUC

Gapop\_60.0 , Gapext 60.0

Searched: 3373863 seqs, 212409041 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1847

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001as.\*
- 5: Geneseq2001bs.\*
- 6: Geneseq2002s.\*
- 7: Geneseq2003as.\*
- 8: Geneseq2003bs.\*
- 9: Geneseq2003cs.\*
- 10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	501	100.0	53789	2 AAV21187	AAV21187 Amycolato
2	26	5.2	14055	7 AAL61170	AAL61170 Actinosyn
3	26	5.2	82746	7 AAL61224	AAL61224 Actinosyn
4	26	5.2	88421	6 AAL40781	AAL40781 842int g
5	22	4.4	9858	8 ACC38255	ACC38255 Simocycli
6	22	4.4	37856	6 AAL11992	AAL11992 S. cellul
7	20	4.0	588	6 ABQ91233	ABQ91233 M. capsul
8	20	4.0	1091	3 AAF12516	AAF12516 Aspergill
9	20	4.0	5275	2 AAV26962	AAV26962 Bovine pa
10	20	4.0	5275	2 AAT95857	AAT95857 Bovine pa
11	20	4.0	5275	2 AAZ25053	AAZ25053 Bovine pa
12	20	4.0	5275	2 AAV82483	AAV82483 Bovine pa
13	20	4.0	5275	3 AAZ89296	AAZ89296 Bovine ca
14	20	4.0	5275	6 AAI72120	AAI72120 cDNA enco
15	19	3.8	558	9 ADC08850	ADC08850 Corn DNA
16	19	3.8	1172	2 AAV69527	AAV69527 Soil bact
17	19	3.8	1178	2 AAV69562	AAV69562 Soil deri
18	19	3.8	1818	7 ACC44559	ACC44559 Maize gra
19	19	3.8	1915	2 AAX60319	AAX60319 DNA seque
20	19	3.8	2223	7 ACC44560	ACC44560 Maize 797
21	19	3.8	2267	2 AAX63355	AAX63355 Granule b
22	19	3.8	3738	7 ACA25719	ACA25719 Prokaryot
23	19	3.8	4800	2 AAV29752	AAV29752 Zea may

c	24	19	3.8	4800	7 ABX09931	Abx09931 DNA encod
c	25	19	3.8	6029	4 AAK82712	AAK82712 Human imm
c	26	19	3.8	21185	3 AAG63350	AAG63350 Streptomy
c	27	19	3.8	24379	2 AAT93095	AAT93095 Streptomy
c	28	19	3.8	24379	2 AAV25925	AAV25925 Streptomy
c	29	19	3.8	63184	3 AAG63346	AAG63346 Streptomy
c	30	19	3.8	349980	6 ABQ81846	ABQ81846 Bifidobac
c	31	18	3.6	175	4 AAI29319	AAI29319 Colon tum
c	32	18	3.6	175	7 ABZ33505	ABZ33505 Human col
c	33	18	3.6	313	6 ABN18020	ABN18020 Human ORF
c	34	18	3.6	418	8 ACH19833	ACH19833 Human adu
c	35	18	3.6	457	5 AAH83034	AAH83034 Human ova
c	36	18	3.6	498	2 AAV29373	AAV29373 Human sma
c	37	18	3.6	537	5 AAS75981	AAS75981 DNA encod
c	38	18	3.6	622	4 AAL01672	AAL01672 Human rep
c	39	18	3.6	622	4 ABA07591	ABA07591 Human ova
c	40	18	3.6	633	5 AAS81267	AAS81267 DNA encod
c	41	18	3.6	819	8 ACC83519	ACC83519 Human exo
c	42	18	3.6	909	7 ACA44265	ACA44265 Prokaryot
c	43	18	3.6	1699	4 AAH14082	AAH14082 Human CDN
c	44	18	3.6	1703	6 ABL90390	ABL90390 Human pol
c	45	18	3.6	1715	2 AAZ24887	AAZ24887 Human sec

#### ALIGNMENTS

RESULT 1  
AAV21187  
ID AAV21187 standard; DNA; 53789 BP.  
XX  
AC AAV21187;  
XX  
DT 24-JUL-1998 (first entry)  
XX  
DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.  
XX  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin; ds.  
XX  
OS Amycolatopsis mediterranei.  
XX  
FH Key Location/Qualifiers  
CDS 1825..15543  
FT /\*tag= a  
FT /label= ORF\_A  
FT /product= "polyketide synthase"  
FT 15550..30759  
FT /\*tag= b  
FT /label= ORF\_B  
FT /product= "polyketide synthase"  
FT 30895..36060  
FT /\*tag= C  
FT /label= ORF\_C  
FT /product= "polyketide synthase"  
FT 36259..41325  
FT /\*tag= d  
FT /label= ORF\_D  
FT /product= "polyketide synthase"  
FT 41373..51614  
FT /\*tag= e  
FT /label= ORF\_E  
FT /product= "polyketide synthase"  
FT 51713..52993  
FT /\*tag= f  
FT /label= ORF\_F  
FT /product= "polyketide synthase"

WO9807868-A1.

26-FEB-1998.

18-AUG-1997; 97WO-EP004495.

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
DR P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX  
XX Claim 4; Page 53-102; 205pp; English.  
XX  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
XX  
XX Query Match 100.0%; Score 501; DB 2; Length 53789;  
XX Best Local Similarity 100.0%; Pred. No. 7.1e-230;  
XX Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 ATGTTCTACAGTCGGGCGACGACCGGGCGGCCCAAGAGGCGTGTCTCCAGCGGCGAAC 60  
XX 1825 ATGTTCTACAGTCGGGCGACGACCGGGCGGCCCAAGAGGCGTGTCTCCAGCGGCGAAC 1884  
XX  
XX 61 TGCCTCTGTCGCTTCTCTAGTCGCTCCCGGTTGTCGGACCGACGCGG 120  
XX 1885 TGCCTCTGTCGCTTCTCTAGTCGCTCCCGGTTGTCGGACCGACGCGG 1944  
XX  
XX 121 GTGCTCTGGCGCTCCCGCTGTCTCCAGAGCTTTTCGACATCGGCTCGTCTCTCGGCC 180  
XX 1945 GTGCTCTGGCGCTCCCGCTGTCTCCAGAGCTTTTCGACATCGGCTCGTCTCTCGGCC 2004  
XX  
XX 181 ACCGTGTCTGGGCGACGCTCGGATCGCGATCGGACGCTCGGCGACGAGTATCGG 240  
XX 2005 ACCGTGTCTGGGCGACGCTCGGATCGCGATCGGACGCTCGGCGACGAGTATCGG 2064  
XX  
XX 241 CTGATCGAGCGGAGAGCTCGACTTCTCGCGGCGTGGCGACCTACACCACTG 300  
XX 2065 CTGATCGAGCGGAGAGCTCGACTTCTCGCGGCGTGGCGACCTACACCACTG 2124  
XX  
XX 301 GTGCGGCGCGCGCGGCGACGCGGTTTCTCGCGCGCGAGCCTCGGATCGGCGG 360  
XX 2125 GTGCGGCGCGCGCGGCGACGCGGTTTCTCGCGCGCGAGCCTCGGATCGGCGG 2184  
XX  
XX 361 GGCGCGGCTCTCGGCGCGGCGTCCGAGAGAGTTCGAGAGAGCTTCGGGTCGCGCTG 420  
XX 2185 GGCGCGGCTCTCGGCGCGGCGTCCGAGAGAGTTCGAGAGAGCTTCGGGTCGCGCTG 2244  
XX  
XX 421 ATCGACCGCTACGCGACGACGAGCTTCGGGCGGATCACCATGAACCGCGGACGCG 480  
XX 2245 ATCGACCGCTACGCGACGACGAGCTTCGGGCGGATCACCATGAACCGCGGACGCG 2304  
XX  
XX 481 GCCCGCGCTCGAGGCGCTCGTGC 501

Db 2305 GCCCGCGTCGAGGGCTCGTGC 2325

## RESULT 2

AAAL61170  
ID AAL61170 standard; DNA; 14055 BP.

XX AAL61170;

XX 22-SEP-2003 (first entry)

XX Actinosynnema pretiosum polyketide synthase (PKS) gene #1.

XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX gene; ds.

XX Actinosynnema pretiosum.

XX WO2003045312-A2.

XX 05-JUN-2003.

XX 21-NOV-2002; 2002WO-US037547.

XX 21-NOV-2001; 2001US-0332158P.

XX (UNIW ) UNIV WASHINGTON.

XX Floss HG, Yu T, Leistner B;

XX WPI; 2003-493374/46.

XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.

XX Disclosure; Page 45-53; 160pp; English.

XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene

XX Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;

Query Match 5.2%; Score 26; DB 7; Length 14055;

Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 ATCGAGCGCTACGCGAGCAGCCGAGAC 446

Db 904 ATCGAGCGCTACGCGAGCAGCCGAGAC 929

## RESULT 3

AAAL61224/c

ID AAL61224 standard; DNA; 82745 BP.

XX AAL61224;

XX 22-SEP-2003 (first entry)

XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.

XX Maytansinoid; ansamitocin; antitumour; gene; ds.

XX Actinosynnema pretiosum.

XX WO2003045312-A2.

XX 05-JUN-2003.



```
XX 21-NOV-2002; 2002WO-US037547.
XX PF
XX PR
XX 21-NOV-2001; 2001US-0332158P.
XX PA (UNIW ) UNIV WASHINGTON.
XX Floss HG, Yu T, Leistner E;
XX WPI; 2003-493374/46.
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Claim 7; Page 105-152; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin ansamitocin biosynthetic gene cluster I
XX
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;
XX
XX Query Match 5.2%; Score 26; DB 7; Length 82746;
XX Best Local Similarity 100.0%; Pred. No. 0.0095;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 421 ATCGAGCGCTACGGCAGCAGCGAGAC 446
Db 36138 ATCGAGCGCTACGGCAGCAGCGAGAC 36113
RESULT 4
AAL40781
ID AAL40781 standard; DNA; 88421 BP.
AC AAL40781;
XX
DT 03-OCT-2002 (first entry)
XX
DE 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.
XX
XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
XX biosynthesis gene cluster; bioengineering; peptide synthetase module;
XX adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;
XX chlorinate; lipdepsipeptide; gene; ds.
XX
XX Actinoplanes sp.
XX
XX Key Location/Qualifiers
FH 2077..3078
FT CDS /*tag= a
FT /*product= "Protein of ORF 1"
FT 3118..4032
FT CDS /*tag= b
FT /*product= "Protein of ORF 2"
FT 4038..5048
FT CDS /*tag= c
FT /*product= "Protein of ORF 3"
FT complement(6665..5814)
FT /*tag= d
FT /*product= "Protein of ORF 4"
FT complement(7703..6693)
FT /*tag= e
FT /*product= "Protein of ORF 5"
FT complement(9464..8130)
FT /*tag= f
FT /*product= "Protein of ORF 6"
FT 9691..10761
FT CDS /*tag= g
FT /*product= "Protein of ORF 7"
FT
FT complement(12751..10829)
FT /*tag= h
FT /*product= "Protein of ORF 8"
FT complement(13617..12802)
FT /*tag= i
FT /*product= "Protein of ORF 9"
FT complement(15203..13614)
FT /*tag= j
FT /*product= "Protein of ORF 10"
FT 15591..15863
FT CDS /*tag= k
FT /*product= "Protein of ORF 11"
FT 15880..19035
FT CDS /*tag= l
FT /*product= "Protein of ORF 12"
FT 19032..39713
FT CDS /*tag= m
FT /*product= "Protein of ORF 13"
FT 39713..65800
FT CDS /*tag= n
FT /*product= "Protein of ORF 14"
FT 65826..66530
FT CDS /*tag= o
FT /*product= "Protein of ORF 15"
FT 66546..67370
FT CDS /*tag= p
FT /*product= "Protein of ORF 16"
FT 67384..70059
FT CDS /*tag= q
FT /*product= "Protein of ORF 17"
FT 70099..70662
FT CDS /*tag= r
FT /*product= "Protein of ORF 18"
FT 70659..71906
FT CDS /*tag= s
FT /*product= "Protein of ORF 19"
FT complement(73439..71964)
FT /*tag= t
FT /*product= "Protein of ORF 20"
FT complement(74216..73563)
FT /*tag= u
FT /*product= "Protein of ORF 21"
FT complement(75424..74213)
FT CDS /*tag= v
FT /*product= "Protein of ORF 22"
FT 75535..76464
FT CDS /*tag= w
FT /*product= "Protein of ORF 23"
FT complement(78110..76449)
FT /*tag= x
FT /*product= "Protein of ORF 24"
FT complement(79864..78107)
FT CDS /*tag= y
FT /*product= "Protein of ORF 25"
FT complement(81624..79861)
FT /*tag= z
FT /*product= "Protein of ORF 26"
FT complement(81909..81682)
FT CDS /*tag= aa
FT /*product= "Protein of ORF 27"
FT complement(82346..82062)
FT /*tag= ab
FT /*product= "Protein of ORF 28"
FT 82587..84446
FT CDS /*tag= ac
FT /*product= "Protein of ORF 29"
FT 84481..85548
FT CDS /*tag= ad
FT /*product= "Protein of ORF 30"
FT 85556..86845
FT CDS /*tag= ae
FT /*product= "Protein of ORF 31"
FT 87372..86803
FT CDS
```

FT           /\*tag= af  
FT           /product= "Protein of ORF 32"  
FT           87494. .88420  
FT           /\*tag= ag  
FT           /product= "Protein of ORF 33"

XX           WO200231155-A2.

XX           18-APR-2002.

XX           15-OCT-2001; 2001WO-CA001462.

XX           13-OCT-2000; 2000US-0239924P.

XX           12-APR-2001; 2001US-0283296P.

XX           24-JUL-2001; 2001US-00910813.

XX           (ECOP-) ECOPTA BIOSCIENCES INC.

XX           Farnet CM, Zazopoulos E, Staffa A;

XX           WPL: 2002-435445/46.

XX           P-PSDB: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,

XX           AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,

XX           AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,

XX           AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,

XX           AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.

XX           Novel isolated ramoplanin biosynthetic pathway polypeptide useful for

XX           chemically modifying biological molecule that is a substrate for a

XX           polypeptide encoded by a ramoplanin biosynthesis gene cluster.

XX           Disclosure; Page 87-135; 212pp; English.

XX           The invention relates to an isolated ramoplanin biosynthetic pathway

XX           polypeptide selected from a polypeptide of open reading frames (ORF) 1-

XX           32. The isolated polypeptides are useful for chemically modifying a

XX           biological molecule that is a substrate for a polypeptide encoded by a

XX           ramoplanin biosynthesis gene cluster, by contacting the biological

XX           molecule with the isolated polypeptide, where the polypeptide chemically

XX           modifies the biological molecule. The method comprises contacting the

XX           biological molecule with at least two different polypeptides encoded by

XX           ramoplanin ORFs 1-31. The polypeptides are useful for directing the

XX           biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated

XX           gene cluster comprising the ORFs is useful as a substrate for

XX           bioengineering of antibiotic structures. An isolated polypeptide or its

XX           encoding nucleic acid sequence is useful for generating derivatives of

XX           ramoplanin, for improving production or for producing variants of other

XX           antibiotics of the peptide class. The isolated polypeptides are useful

XX           for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain

XX           in conjunction with other peptide synthetase modules and allowing the

XX           incorporation of Thr into a peptide antibiotic precursor, for modifying the

XX           fatty acid structure and/or enhancing fatty acid incorporation into the

FT           RESULT 5  
FT           ACC58255  
FT           ID   ACC58255 standard; DNA; 9858 BP.  
FT           XX  
FT           AC   ACC58255;  
FT           XX  
FT           DT   26-AUG-2003 (first entry)  
FT           XX  
FT           DE   Simocyclinone biosynthetic gene cluster.  
FT           XX  
FT           KW   Simocyclinone; aminocoumarin; antibiotic; antibacterial; cytostatic;  
FT           XX   gene; ds.  
FT           XX  
FT           OS   Streptomyces antibioticus.  
FT           XX  
FT           FH   Key           Location/Qualifiers  
FT           FT   CDS           100..810  
FT           FT           /\*tag= a  
FT           FT           /gene= "simK"  
FT           FT           /product= "SimK"  
FT           FT           /note= "encodes ABR42548, putative reductase"  
FT           FT   CDS           847..2406  
FT           FT           /\*tag= b  
FT           FT           /gene= "simL"  
FT           FT           /product= "SimL"  
FT           FT           /note= "encodes ABR42549, putative amide synthetase"  
FT           FT   CDS           2435..5427  
FT           FT           /\*tag= c  
FT           FT           /gene= "simH"  
FT           FT           /product= "SimH"  
FT           FT           /note= "encodes ABR42550, peptide synthetase-like  
FT           FT   CDS           5440..6690  
FT           FT           /\*tag= d  
FT           FT           /gene= "simI"  
FT           FT           /product= "SimI"  
FT           FT           /note= "encodes ABR42551, cytochrome P450"  
FT           FT   CDS           6727..6939  
FT           FT           /\*tag= e  
FT           FT           /gene= "simY"  
FT           FT           /product= "SimY"  
FT           FT           /note= "encodes ABR42552, mbtH-like protein"  
FT           FT   CDS           6950..7771  
FT           FT           /\*tag= f  
FT           FT           /gene= "simJ1"  
FT           FT           /product= "SimJ1"  
FT           FT           /note= "encodes ABR42553, putative 3-oxoacyl-(ACP)-  
FT           FT   CDS           7941..8696  
FT           FT           /\*tag= g  
FT           FT           /gene= "simI"  
FT           FT           /product= "SimI"  
FT           FT           /note= "encodes ABR42554, putative response regulator"  
FT           FT   CDS           9018..9758  
FT           FT           /\*tag= h  
FT           FT           /gene= "simJ2"  
FT           FT           /product= "SimJ2"  
FT           FT           /note= "encodes ABR42555, putative 3-oxoacyl (ACP)-  
FT           FT   CDS           reductase"  
FT           XX  
FT           PN   WO2003014352-A2.  
FT           XX  
FT           PD   20-FEB-2003.  
FT           XX  
FT           PF   06-AUG-2002; 2002WO-EP008777.  
FT           XX  
FT           PR   08-AUG-2001; 2001US-0310808P.  
FT           XX  
FT           PA   (UYTU-) UNIV TUEBINGEN.  
FT           XX  
FT           PI   Heide I, Li S;  
FT           XX

DR WPI; 2003-268202/26.  
 DR P-PSDB; ABR42548, ABR42549, ABR42550, ABR42551, ABR42552, ABR42553,  
 DR ABR42554, ABR42555.  
 XX  
 PT Novel polypeptide encoded by open reading frame of aminocoumarin  
 PT biosynthetic gene cluster, useful for synthesizing aminocoumarins useful  
 PT for treating bacterial infections and malignancies.  
 XX  
 PS Claim 2; Fig 3; 124pp; English.  
 XX  
 CC The present sequence is a partial sequence of the simocyclinone  
 CC biosynthetic gene cluster of Streptomyces antibioticus. It includes open  
 CC reading frames (ORFs) that show sequence similarity to ORFs of the  
 CC novobiocin biosynthetic gene cluster. The invention provides  
 CC aminocoumarin (e.g. simocyclinone) biosynthetic nucleic acids and  
 CC polypeptides, a method for producing hybrid antibiotics using a  
 CC combination of various regions of biosynthetic gene clusters, and a  
 CC pharmaceutical composition containing an aminocoumarin compound for use  
 CC in the treatment of Gram-positive bacterial infections and of malignant  
 CC diseases  
 XX  
 SQ Sequence 9858 BP; 1512 A; 3292 C; 3482 G; 1572 T; 0 U; 0 Other;  
 Query Match 4.4%; Score 22; DB 8; Length 9858;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 CGACCGGGCGGCCCAAGGGCGT 41  
 Db 4134 CGACCGGGCGGCCCAAGGGCGT 4155  
 RESULT 6  
 AA11992  
 ID AA11992 standard; DNA; 37856 BP.  
 AC AA11992;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 07-AUG-2000 (first entry)  
 XX  
 DE S. cellulose DNA encoding polyketide and heteropolyketide enzymes.  
 XX  
 XX Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;  
 KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;  
 KW plant-protection; ds.  
 XX  
 OS Polyangium cellulose.  
 XX  
 FH Location/Qualifiers  
 CDS complement(3398..6100)  
 FT /\*tag= a  
 FT /product= "ORF1-tRNA synthetase"  
 FT /note= "gtg start codon"  
 FT complement(6374..7111)  
 FT /\*tag= b  
 FT /product= "ORF2-monoxygenase"  
 FT complement(8433..9550)  
 FT /\*tag= c  
 FT /product= "ORF3-aminotransferase"  
 FT /note= "AGT start codon given in the specification"  
 FT 9855..11393  
 FT /\*tag= d  
 FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"  
 FT /note= "GTG start codon"  
 FT 12212..13658  
 FT /\*tag= e  
 FT /product= "ORF5-3-oxoacyl-ACP-reductase"  
 FT /note= "ACC start codon"  
 FT 15374..19984  
 FT /\*tag= f  
 FT /product= "ORF6-polyketide synthase"  
 FT 20003..27889

FT /\*tag= g  
 FT /product= "ORF7-peptide synthetase"  
 FT 28251..29400  
 FT /\*tag= h  
 FT /product= "ORF8-transpeptidase"  
 FT complement(30040..31720)  
 FT /\*tag= i  
 FT /product= "ORF9-regulation element"  
 FT /note= "CGC stop codon"  
 FT 31982..32932  
 FT /\*tag= j  
 FT /product= "ORF10-transcription regulator"  
 FT 33128..33613  
 FT /\*tag= k  
 FT /product= "ORF11-regulation element"  
 FT /note= "GTG start codon"  
 FT 33661..34077  
 FT /\*tag= l  
 FT /product= "ORF12-regulation element"  
 FT complement(35255..35616)  
 FT /\*tag= m  
 FT /product= "ORF13-transcription regulator"  
 FT complement(35730..36242)  
 FT /\*tag= n  
 FT /product= "ORF14-transcription regulator"  
 FT /note= "GTG start codon"  
 XX  
 PN DE19846493-A1.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 09-OCT-1998; 98DE-01046493.  
 XX  
 XX 09-OCT-1998; 98DE-01046493.  
 XX  
 XX (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
 XX Beyer S, Mueller R;  
 DR WPI; 2000-294101/26.  
 XX  
 PT DNA sequence coding for products involved in the biosynthesis of  
 PT polyketide or heteropolyketide compounds, especially epothilone.  
 XX  
 PS Claim 3; Page 20-33; 36pp; German.  
 XX  
 CC This invention describes a novel DNA sequence (I) whose expression  
 CC products effect or are involved in the enzymatic biosynthesis,  
 CC mutasynthesis or partial synthesis of polyketide or heteropolyketide  
 CC compounds (II). (I) can be inserted into an expression vector and used to  
 CC transform or transfect prokaryotic or eukaryotic cells with the aim of  
 CC obtaining strains that produce large amounts of polyketide or  
 CC heteropolyketide compounds, especially epothilones, which have cytotoxic  
 CC and/or immunosuppressant and antibiotic and antifungal activities and are  
 CC useful as plant-protection agents. This sequence represents the DNA  
 CC sequence isolated from Sorangium cellulosum which is described in the  
 CC method of the invention. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 U; 0 Other;  
 Query Match 4.4%; Score 22; DB 3; Length 37856;  
 Best Local Similarity 100.0%; Pred. No. 0.85;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 CGACCGGGCGGCCCAAGGGCGT 41  
 Db 21906 CGACCGGGCGGCCCAAGGGCGT 21927  
 RESULT 7  
 ABQ91233  
 ID ABQ91233 standard; DNA; 588 BP.  
 XX

```

AC ABQ91233;
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #1218 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
FN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002WO-N0000019.
XX
PR 12-JAN-2001; 2001NO-00000235.
XX
PR 12-JAN-2001; 2001NO-00000239.
XX
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557818/59.
XX
XX Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes.
XX
PS Claim 14; Page 502; 678pp; English.
XX
XX The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC invention
XX
SQ Sequence 588 BP; 105 A; 196 C; 196 G; 91 T; 0 U; 0 Other;

Query Match 4.0%; Score 20; DB 6; Length 588;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 CGGCGTGCCGACCACTACC 292
Db 447 CGGCGTGCCGACCACTACC 466

RESULT 8
AAFL12516/c
ID AAF12516 standard; cDNA; 1091 BP.
XX
AC AAF12516;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5039.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
FN WO200056762-A2.
XX

```

---

```

XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US007781.
XX
XX 22-MAR-1999; 99US-00273623.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags.
XX
XX Claim 88; Page 2116; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring the
XX global expression of genes from FF cells allows the production potential
XX of the microorganisms to be improved. New genes may be discovered,
XX possible functions of unknown open reading frames can be identified and
XX gene copy number variation and stability can be monitored. The expression
XX of genes can be used to study how FF cells adapt to changes in culture
XX conditions, environmental stress, spore morphogenesis, recombination,
XX metabolic or catabolic pathway engineering. Using ESTs provides several
XX advantages over genomic or random cDNA clones including elimination of
XX redundancy as one spot on an array equals one gene or open reading frame,
XX and organisation of the microarrays based on function of the gene
XX products to facilitate analysis of the results. AAF07478 to AAF11247
XX represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents
XX ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
XX Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from
XX Trichoderma reesei, which are all specifically claimed in the present
XX invention
XX
SQ Sequence 1091 BP; 240 A; 305 C; 308 G; 238 T; 0 U; 0 Other;

Query Match 4.0%; Score 20; DB 3; Length 1091;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 CGGCCGCGCAGCGCGTTTC 327
Db 656 CGGCCGCGCAGCGCGTTTC 637

RESULT 9
AAV26962
ID AAV26962 standard; cDNA to mRNA; 5275 BP.
XX
AC AAV26962;
XX
DT 01-SEP-1998 (first entry)
XX
DE Bovine parathyroid calcium receptor 1 gene 5kb fragment.
XX
KW ss; calcium ion concentration; parathyroid hormone; homeostasis; kidney;
KW calcium receptor; detection.
XX
OS Bos sp.
XX
XX Key Location/Qualifiers
FH 515. .3772
FT CDS

```



```

XX 05-OCT-1999.
XX 03-OCT-1997; 97US-00943986.
XX 23-FEB-1993; 93US-00009389.
XX 22-OCT-1993; 93US-00141248.
XX 19-AUG-1994; 94US-00292827.
XX 21-OCT-1994; 94WO-US012117.
XX 08-DEC-1994; 94US-00353784.
XX 07-JUN-1995; 95US-00484565.
XX (NPSP-) NPS PHARM INC.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX Brown EM, Hebert SC, Garrett JE;
XX WPI; 1999-571274/48.
XX P-PSDB; AAY41778.
XX Nucleic acids encoding protein calcium receptors useful for identifying
XX modulators of receptor expression and activity and for the production of
XX antigens specific for calcium receptors.
XX Claim 5; Fig 47; 174pp; English.
XX The present sequence encodes bovine parathyroid calcium receptor 1
XX (BoPCar 1). Calcium receptor polynucleotides may be used: (i) for
XX producing receptor proteins (or fragments) useful for determining
XX structure and activity relationships, for assaying molecular activity on
XX the receptor (i.e. identifying modulators of receptor function) and for
XX producing antibodies specific for the receptor; (ii) for sequencing the
XX normal form of the nucleic acids (the derived sequence may be compared
XX with other receptors to identify conserved sequences, mutations and
XX variations that may influence calcium receptor activity and to determine
XX target sites for antisense molecules, ribozymes, hybridization probes and
XX polymerase chain reaction (PCR) amplification primers; (iii) as
XX hybridization probes to detect the presence of similar sequences in
XX samples; and (iv) as PCR primers to generate particular nucleic acid
XX sequence regions; (e.g. to generate regions to be probed by hybridization
XX detection probes)
XX Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;
XX
XX Query Match 4.0%; Score 20; DB 2; Length 5275;
XX Best Local Similarity 100.0%; Pred. No. 9.8;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 362 GCGGGGTCTCGGCGCGGG 381
Db 223 GCGGGGTCTCGGCGCGGG 242
XX
RESULT 12
AAV82483
ID AAV82483 standard; cDNA to mRNA; 5275 BP.
XX
XX AAV82483;
XX
XX 19-MAR-1999 (first entry)
XX
XX Bovine parathyroid calcium receptor BoPCar 1 encoding cDNA.
XX
XX Parathyroid calcium receptor; inorganic ion receptor; osteoporosis;
XX calcium homeostasis; hyperparathyroidism; seizure; stroke; epilepsy;
XX spinal cord injury; hypoxia-induced nerve cell damage; cardiac arrest;
XX neonatal distress; neurodegenerative disease; Alzheimer's disease;
XX Huntington's disease; Parkinson's disease; dementia; muscle tension;
XX depression; anxiety; ss.
XX
XX Bos sp.
XX
XX Key Location/Qualifiers

```

```

FT CDS 515..3772
FT /*tag= a
XX
XX US5858684-A.
XX 12-JAN-1999.
XX 07-JUN-1995; 95US-00480751.
XX 23-AUG-1991; 91US-00749451.
XX 11-FEB-1992; 92US-00834044.
XX 21-AUG-1992; 92US-00934161.
XX 12-FEB-1993; 93US-00017127.
XX 23-FEB-1993; 93US-00009389.
XX 22-OCT-1993; 93US-00141248.
XX 19-AUG-1994; 94US-00292827.
XX 21-OCT-1994; 94WO-US012117.
XX 08-DEC-1994; 94US-00353784.
XX (NPSP-) NPS PHARM INC.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX Del Mar EG, Balandrin MF, Van Wagenen BC, Nemeth EF, Brown EM;
XX Garrett JE, Hebert SC;
XX WPI; 1999-119871/10.
XX P-PSDB; AAW89563.
XX
XX Screening for calcium receptor-active compounds - by recombinant
XX expression of nucleic acid encoding calcium receptor and determining the
XX effect of compounds on calcium receptor activity.
XX Claim 1; Fig 47; 176pp; English.
XX A method has been developed of screening for a compound able to affect
XX one or more activities of a calcium receptor (CR) comprises: (A)
XX contacting a recombinant cell with a test compound, where the recombinant
XX cell comprises a recombinant nucleic acid expressing the CR, provided
XX that the cell does not have functional CR expression from endogenous
XX nucleic acid; (B) determining the ability of the test compound to affect
XX one or more activities of the calcium receptor; and (C) comparing the
XX CR activities in a cell not comprising the recombinant nucleic acid. The
XX present sequence encodes bovine parathyroid CR, designated a BoPCar 1.
XX The nucleic acid sequence of BoPCar 1 can be used as part of the
XX recombinant nucleic acid in the method described above. The compounds
XX identified can be used to treat diseases or disorders characterised by
XX abnormal calcium homeostasis, e.g. hyperparathyroidism, osteoporosis and
XX other bone and mineral-related disorders. They can also be used for the
XX treatment of diseases and disorders associated with disrupted Ca2+
XX responses, e.g. seizures, stroke, spinal cord injury, hypoxia-induced
XX nerve cell damage such as in cardiac arrest or neonatal distress,
XX epilepsy, neurodegenerative diseases such as Alzheimer's disease,
XX Huntington's disease and Parkinson's disease, dementia, muscle tension,
XX depression, and anxiety
XX Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;
XX
XX Query Match 4.0%; Score 20; DB 2; Length 5275;
XX Best Local Similarity 100.0%; Pred. No. 9.8;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 362 GCGGGGTCTCGGCGCGGG 381
Db 223 GCGGGGTCTCGGCGCGGG 242
XX
RESULT 13
AAZ82296
ID AAZ82296 standard; cDNA to mRNA; 5275 BP.
XX
XX AAZ82296;
XX

```

DT 09-JUN-2000 (first entry)  
DE Bovine calcium receptor BoPCar1 cDNA.  
XX  
KW Calcium receptor; treatment; calcimimetic; calcilytic; osteopathic;  
KW cerebroprotective; cytosstatic; neuroprotective; dermatological;  
KW tranquilizer; vulnerar; antitumor; immunosuppressive; hypotensive;  
KW cardiant; parathyroid hormone; osteoporosis; calcitonin secretion;  
KW hyperparathyroidism; Paget's disease; bovine; ss.  
XX  
OS Bos taurus.  
XX  
XX Key Location/Qualifiers  
FT 515..3772  
FT CDS /\*tag= a  
FT /product= "calcium receptor"  
XX  
PN US6031003-A.  
XX  
PD 29-FEB-2000.  
XX  
XX 07-JUN-1995; 95US-00484719.  
XX  
XX 23-AUG-1991; 91US-00749451.  
PR 11-FEB-1992; 92US-00834044.  
PR 21-AUG-1992; 92US-00934161.  
PR 12-FEB-1993; 93US-00017127.  
PR 23-FEB-1993; 93US-00009389.  
PR 22-OCT-1993; 93US-00141248.  
PR 19-AUG-1994; 94US-00292827.  
PR 21-OCT-1994; 94WO-US012117.  
PR 08-DEC-1994; 94US-00353784.  
XX  
XX (NPSP-) NPS PHARM INC.  
XX (BGHM) BRIGHAM & WOMENS HOSPITAL.  
XX  
PI Balandrin MF, Delmar EG, Moe ST, Nemeth EF, Van Wagenen BC;  
XX  
XX WPI; 2000-301969/26.  
DR P-PSDB; AAB51825.  
XX  
XX Treating disorders or diseases in a patient by modulating inorganic ion  
PT receptor activities especially calcium receptor for hyperparathyroidism,  
PT by administering a calcimimetic or calcilytic compound.  
XX  
XX Example 25; Col 107-116; 194pp; English.  
XX  
XX This invention describes a novel method of treating disorders by  
CC modulating calcium receptor activity in vitro by administering a  
CC calcimimetic, or a calcilytic compound. The products of the invention  
CC have osteopathic, cerebroprotective, cytosstatic, neuroprotective,  
CC dermatological, tranquilizer, vulnerar, antitumor, immunosuppressive,  
CC hypotensive and cardiant activity. The method is suitable for reducing  
CC parathyroid hormone level in a patient to that of a normal individual,  
CC treating a patient having osteoporosis, to inhibit bone resorption, and  
CC to stimulate calcitonin secretion in vitro or in vivo. The level of  
CC parathyroid hormone is reduced to cause a decrease in plasma Ca2+. The  
CC method is useful in treating disorders in humans such as  
CC hyperparathyroidism, Paget's disease and osteoporosis. Also for treatment  
CC or prevention, based on the affected cells, of other disorders and  
CC conditions like seizures, stroke, head trauma, spinal cord injury,  
CC hypoxia-induced nerve cell damage such as in cardiac arrest or neonatal  
CC distress, epilepsy, Alzheimer's disease, Huntington's disease,  
CC Parkinson's disease, dementia, muscle tension, depression, anxiety, panic  
CC disorder, OCD (not defined), post-traumatic stress disorder,  
CC schizophrenia, neuroleptic malignant syndrome and Tourette's syndrome,  
CC diseases involving excess water reabsorption by the kidney such as  
CC syndrome of inappropriate ADH secretion (SIADH), cirrhosis, congestive  
CC heart failure and nephrosis, hypertension, preventing and/or decreasing  
CC renal toxicity from cationic antibiotics (e.g. aminoglycoside  
CC antibiotics), gut motility disorders such as diarrhoea and spastic colon,  
CC GI (gastrointestinal) ulcer diseases, GI diseases with excessive calcium  
CC absorption such as sarcoidosis and autoimmune diseases and organ

CC transplant rejection. This sequence encodes the bovine calcium receptor  
CC BoPCar1 which is described in the method of the invention  
XX  
SQ Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;  
Query Match 4.0%; Score 20; DB 3; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 362 GCGCGTCTCGCGCGCGG 381  
DB 223 GCGCGTCTCGCGCGCGG 242  
RESULT 14  
AAI72120  
ID AAI72120 standard; cDNA to mRNA; 5275 BP.  
XX  
XX AAI72120;  
XX  
XX 07-AUG-2003 (revised)  
DT 25-MAR-2002 (first entry)  
XX  
XX cDNA encoding BoPCar1.  
XX  
XX Calcium receptor; bovine; human; parathyroid; calcium receptor; thyroid;  
XX C-cell; inorganic ion receptor; homeostasis; sarcoidosis;  
XX hyperparathyroidism; osteoporosis; central nervous system; seizure;  
XX stroke; head trauma; spinal cord injury; organ transplant rejection;  
XX hypoxia-induced nerve cell damage; cardiac arrest; neonatal distress;  
XX epilepsy; neurodegenerative disease; Alzheimer's disease; cirrhosis;  
XX Huntington's disease; Parkinson's disease; dementia; depression; anxiety;  
XX panic disorder; obsessive-compulsive disorder; spastic colon;  
XX post-traumatic stress disorder; schizophrenia; diarrhoea; kidney;  
XX neuroleptic malignant syndrome; Tourette's syndrome; gut motility;  
XX inappropriate ADH secretion; SIADH; gastrointestinal ulcer disease;  
XX congestive heart failure; nephrosis; hypertension;  
XX aminoglycoside antibiotic; ss.  
XX  
XX Bos taurus.  
XX  
XX Key Location/Qualifiers  
FT 515..3772  
FT CDS /\*tag= a  
FT /product= "BoPCar1"  
XX  
XX US6313146-B1.  
XX  
XX 06-NOV-2001.  
XX  
XX 07-JUN-1995; 95US-00484159.  
XX  
XX 23-AUG-1991; 91US-00749451.  
PR 11-FEB-1992; 92US-00834044.  
PR 21-AUG-1992; 92US-00934161.  
PR 12-FEB-1993; 93US-00017127.  
PR 23-FEB-1993; 93US-00009389.  
PR 22-OCT-1993; 93US-00141248.  
PR 19-AUG-1994; 94US-00292827.  
PR 21-OCT-1994; 94WO-US012117.  
PR 08-DEC-1994; 94US-00353784.  
XX  
XX (NPSP-) NPS PHARM INC.  
XX  
XX Van Wagenen BC, Balandrin MF, Delmar EG, Nemeth EF;  
XX WPI; 2002-081872/11.  
DR P-PSDB; AAB47820.  
XX  
XX Novel inorganic ion receptor-modulating compounds, useful for treating  
PT e.g. hyperparathyroidism, osteoporosis, stroke, epilepsy, Alzheimer's  
PT disease, dementia, depression, anxiety, hypertension, cirrhosis and  
PT spastic colon.

XX Example 25; Fig 47; 227pp; English.

XX The sequences given in A172120-23 encode various calcium receptor

XX proteins. This sequence encodes bovine parathyroid calcium receptor. The

XX calcium receptor proteins are used, in conjunction with the compounds of

XX the invention, for structure determination, to assay a molecule's activity

XX on a receptor and to obtain antibodies to that receptor. The compounds of

XX the invention, which modulate inorganic ion receptors are useful for

XX treating and diagnosing diseases or disorders which can be treated by

XX modulating inorganic ion receptor activity. This is preferably a disease

XX or disorder characterized by abnormal inorganic ion homeostasis,

XX preferably abnormal calcium homeostasis (hyperparathyroidism,

XX osteoporosis and other bone and mineral-related disorders), an abnormal

XX level of a messenger whose production or secretion is affected by

XX inorganic ion receptor activity, and an abnormal level or activity of a

XX messenger whose function is affected by inorganic ion receptor activity.

XX These compounds are also useful for treating and diagnosing diseases or

XX disorders of the central nervous system such as seizures, stroke, head

XX trauma, spinal cord injury, hypoxia-induced nerve cell damage such as in

XX cardiac arrest or neonatal distress, epilepsy, neurodegenerative diseases

XX such as Alzheimer's disease, Huntington's disease and Parkinson's

XX disease, dementia, depression, anxiety, panic disorder, obsessive-

XX compulsive disorder, post-traumatic stress disorder, schizophrenia,

XX neuroleptic malignant syndrome and Tourette's syndrome, diseases

XX involving excess water reabsorption by the kidney such as inappropriate

XX ADH secretion (SIADH), cirrhosis, congestive heart failure, nephrosis,

XX hypertension, for preventing and/or decreasing renal toxicity from

XX cationic antibiotics (e.g. aminoglycoside antibiotics), gut motility

XX disorders such as diarrhoea, and spastic colon, gastrointestinal (GI)

XX ulcer diseases, GI diseases with excessive calcium absorption such as

XX sarcoidosis, and autoimmune diseases and organ transplant rejection.

XX (Updated on 07-AUG-2003 to correct OS field.)

XX SQ Sequence 5275 BP; 1277 A; 1475 C; 1316 G; 1207 T; 0 U; 0 Other;

Query Match 4.0%; Score 20; DB 6; Length 5275;

Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGTCTCGCGCGCGG 381

DB 223 GCGCGTCTCGCGCGCGG 242

RESULT 15

ADC08850/c

ID ADC08850 standard; DNA; 558 BP.

AC ADC08850;

DT 18-DEC-2003 (first entry)

XX Corn DNA sequence Seq ID1155 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;

XX carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;

XX tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;

XX wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;

XX gene; ds; plant.

XX Zea mays.

OS

XX WO200300905-A2.

PN 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

PF

XX 22-JUN-2001; 2001US-0300112P.

PR 26-SEP-2001; 2001US-032527P.

PR 20-DEC-2001; 2001US-0342327P.

XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;

XX WPI; 2003-229341/22.

XX New plant genes encoding polypeptides having an activity involved in or

XX associated with the synthesis, metabolism or degradation of carbohydrates

XX in the plant grain useful in generating plants having improved

XX nutritional properties.

PS Disclosure; SEQ ID NO 1155; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel

XX polynucleotides comprising a nucleotide sequence encoding a protein which

XX is involved in or associated with the synthesis, metabolism or

XX degradation of carbohydrates in the plant grain and the expression of

XX which is up-regulated during grain filling. The plant is selected from

XX corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,

XX sugarbeet, wheat, and rice. The invention may be useful for the

XX improvement of protein, oil, starch, fibre and moisture content of the

XX cereal grains. In addition, carbohydrate levels may be modified to a more

XX desirable level using the present invention. The present sequence is a

XX DNA sequence from corn which showed homology to rice "grain filling"

XX genes of the invention. Note: The sequence data for this patent did not

XX form part of the printed specification, but was obtained in electronic

XX format directly from WIPO at ftp.wipo.int/pub/publishedpot\_sequences.

XX SQ Sequence 558 BP; 108 A; 173 C; 198 G; 79 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 9; Length 558;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGCGGTGCGGACCACT 289

DB 395 GCGCGGTGCGGACCACT 377

RESULT 16

AAV69527/c

ID AAV69527 standard; DNA; 1172 BP.

XX AAV69527;

XX 15-MAR-1999 (first entry)

XX Soil bacteria peptide synthase clone ps32 DNA.

XX Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;

XX therapeutic; immunosuppressor; antitumour agent; pathogen;

XX genetic diversity; ss.

XX Bacteria.

XX WO9853057-A2.

PN 26-NOV-1998.

XX 21-MAY-1998; 98WO-CA000488.

PF 22-MAY-1997; 97US-00861774.

PR (TERR-) TERRAGEN DIVERSITY INC.

XX Waters B, Miao VEW, Yap WH, Seow KT;

PI WPI; 1999-070158/06.

DR P-FSDB; AAW82675.

XX New degenerate primers - used for recovering antibiotic biosynthetic DNA

XX from soil/lichen material.



XX PS Claim 17; Page 25; 98pp; English.

XX CC This sequence encodes a peptide synthase from soil/lichen material. The

XX CC encoded protein is used in a method for the recovery of antibiotic

XX CC biosynthetic DNA from humic materials or lichen. The PCR products of the

XX CC invention have the potential to be used as therapeutic molecules

XX CC including antibiotics, immunosuppressors and antitumour agents. The

XX CC method allows access to the reservoir of genetic diversity in soil

XX CC pathogenic micro-organisms, in order to find new antibiotics. It also

XX CC allows access to novel biosynthetic genes/enzymes that can be used to

XX CC produce antibiotics or produce specific compounds, enzymatically, in

XX CC vitro

XX SQ Sequence 1172 BP; 209 A; 388 C; 398 G; 177 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1172;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCGGCAGCAGC 23

Db 1171 TCTACACGTCGGCAGCAGC 1153

RESULT 17

AAV69562/c

ID AAV69562 standard; DNA; 1178 BP.

XX AC AAV69562;

XX CC

DT 17-OCT-2003 (revised)

DT 15-MAR-1999 (first entry)

XX CC

DE Soil derived peptide synthase clone ps25 DNA.

XX KW Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;

XX KW therapeutic; immunosuppressor; antitumour agent; pathogen;

XX KW genetic diversity; ss.

XX CC unidentified.

XX CC

PN WC9853097-A2.

XX CC

PD 26-NOV-1998.

XX CC

PF 21-MAY-1998; 98WO-CA000488.

XX CC

PR 22-MAY-1997; 97US-00861774.

XX CC

XX (TERR-) TERRAGEN DIVERSITY INC.

PA Waters B, Miao VFW, Yap WH, Seow KT;

PI WPI; 1999-070158/06.

DR P-PSDB; AAW62714.

XX CC

XX New degenerate primers - used for recovering antibiotic biosynthetic DNA

PT from soil/lichen material.

PS Claim 17; Page 89-90; 98pp; English.

XX CC This sequence encodes a peptide synthase clone, isolated from soil. This

XX CC protein is used in a method for the recovery of antibiotic biosynthetic

XX CC DNA from humic materials or lichen. The PCR products of the invention

XX CC have the potential to be used as therapeutic molecules including

XX CC antibiotics, immunosuppressors and antitumour agents. The method allows

XX CC access to the reservoir of genetic diversity in soil pathogenic micro-

XX CC organisms, in order to find new antibiotics. It also allows access to

XX CC novel biosynthetic genes/enzymes that can be used to produce antibiotics

XX CC or produce specific compounds, enzymatically, in vitro. (Updated on 17-

XX CC OCT-2003 to standardise OS field)

SQ Sequence 1178 BP; 184 A; 407 C; 397 G; 190 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1178;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCGGCAGCAGC 23

Db 1177 TCTACACGTCGGCAGCAGC 1159

RESULT 18

ACC44559/c

ID ACC44559 standard; cDNA; 1818 BP.

XX AC ACC44559;

XX CC

DT 02-JUN-2003 (first entry)

XX CC

DE Maize granule-bound starch synthase encoding cDNA SEQ ID NO:7.

XX KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;

XX KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;

XX KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;

XX KW maltodextrin; ethanol; fermentation; beverage; enzyme; gens; ss.

XX CC Zea mays.

XX CC

FH Key Location/Qualifiers

FT CDS 1..1818

FT /\*tag= a

FT /partial

FT /product= "granule-bound starch synthase (waxy)"

FT /note= "no stop codon given"

XX CC

PN WO2003018766-A2.

XX CC

XX 06-MAR-2003.

XX CC

PF 27-AUG-2002; 2002WO-US027129.

XX CC

PR 27-AUG-2001; 2001US-0315281P.

XX CC

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX CC

PI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;

XX CC

DR WPI; 2003-268420/26.

DR P-PSDB; ABP96605.

XX CC

PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.

PT alpha-amylase, useful for producing plant to produce food products having

PT improved taste or fermentable substrates for ethanol.

XX CC

PS Example 2; Page 90-91; 158pp; English.

XX CC

XX The present invention describes polynucleotides which encode processing

XX CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose

XX CC isomerase, or glucoamylase) that are optimised for expression in plants.

XX CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic

XX CC processing enzymes, which are activated under suitable conditions to act

XX CC upon the desired substrate. Also described are self-processing transgenic

XX CC plants and plant parts, e.g. grain, which express one or more of these

XX CC enzymes and have an altered composition that facilitates plant and grain

XX CC processing. Also described is a method (M) for converting starch to

XX CC starch-derived products in a transformed plant part (TPP), by activating

XX CC the starch processing enzyme contained in it. Transgenic grain is useful

XX CC for preparing maltodextrin. A transformed plant (TP) can be used to

XX CC produce food products having improved taste and to produce fermentable

XX CC substrates for ethanol and fermented beverages. (M) eliminates the need

XX CC to mill or physically disrupt the integrity of plant parts prior to

XX CC recovery of starch-derived products. The present sequence encodes maize

XX CC granule-bound starch synthase (waxy), which is given in the

```
CC exemplification of the present invention
XX
SQ Sequence 1818 BP; 346 A; 596 C; 609 G; 267 T; 0 U; 0 Other;

Query Match      3.8%; Score 19; DB 7; Length 1818;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289
Db 1655 GCCGCGGTGCCGACCACT 1637

RESULT 19
AAX60319/c
ID AAX60319 standard; DNA; 1915 BP.
XX
AC AAX60319;
XX
DT 23-AUG-1999 (first entry)
XX
DE DNA sequence of the maize waxy gene.
XX
KW Non-glycogen-like polysaccharide production; fermentation; waxy gene;
KW starch synthesis enzyme; glycogen-synthesis enzyme; glycogen synthesis;
KW non-starch branching gene; amylopectin; amylose; plant-like starch; ss.
XX
OS Zea mays.
XX
PN WO9844780-A1.
XX
PD 15-OCT-1998.
XX
PF 03-APR-1998; 98WO-US006660.
XX
PR 04-APR-1997; 97US-0042939P.
XX
PA (EXSE-) EXSEED GENETICS LLC.
XX
PI Guan H, Keeling PL;
XX
DR WPI; 1998-568285/48.
XX
DR P-PSDB; AAY16604.
XX
PT Producing non-glycogen-like polysaccharides in bacteria, fungi or plants
PT - transformed with genes for enzymes involved in starch or glycogen
PT synthesis allows fermentative production of starches with engineered
PT properties.
XX
PS Disclosure; Fig 49; 150pp; English.
XX
CC The specification describes a method for the production of non-glycogen-
CC like polysaccharides in a host. The method comprises transforming a host,
CC suitable for fermentation, with genes encoding starch- or glycogen-
CC synthesis enzymes, and fermenting the transformants. The specification
CC also describes hosts transformed with a gene active in glycogen synthesis
CC and at least one non-starch branching gene, involved in production of
CC amylopectin or amylose in its original host. The method is used to
CC produce plant-like starches by fermentation and new starches in plants.
CC These starches are useful for all food and non-food applications of
CC starch. The present sequence is used in the course of the invention
XX
SQ Sequence 1915 BP; 365 A; 610 C; 641 G; 299 T; 0 U; 0 Other;

Query Match      3.8%; Score 19; DB 2; Length 1915;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289
Db 1652 GCCGCGGTGCCGACCACT 1634

RESULT 20
ACC44560/c
ID ACC44560 standard; cDNA; 2223 BP.
XX
AC ACC44560;
XX
DT 02-JUN-2003 (first entry)
XX
DE Maize 797GL3 granule-bound starch synthase encoding cDNA SEQ ID NO.9.
XX
KW Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
KW pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;
KW mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
KW maltodextrin; ethanol; fermentation; beverage; enzyme; gene; ss.
XX
OS Zea mays.
XX
Key Location/Qualifiers
FH 1..2223
CDS /tag= a
ET /partial
FT /product= "797GL3 granule-bound starch synthase (waxy)"
FT /note= "no stop codon given"
XX
WO2003018766-A2.
XX
PD 06-MAR-2003.
XX
PF 27-AUG-2002; 2002WO-US027129.
XX
PR 27-AUG-2001; 2001US-0315281P.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;
XX
DR WPI; 2003-268420/26.
XX
DR P-PSDB; ABP96606.
XX
PT Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.
PT alpha-amylase, useful for producing plant to produce food products having
PT improved taste or fermentable substrates for ethanol.
XX
PS Claim 1; Page 92-93; 158pp; English.
XX
CC The present invention describes polynucleotides which encode processing
CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose
CC isomerase, or glucoamylase) that are optimised for expression in plants.
CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic
CC processing enzymes, which are activated under suitable conditions to act
CC upon the desired substrate. Also described are self-processing transgenic
CC plants and plant parts, e.g. grain, which express one or more of these
CC enzymes and have an altered composition that facilitates plant and grain
CC processing. Also described is a method (M) for converting starch to
CC starch-derived products in a transformed plant part (TPP), by activating
CC the starch processing enzyme contained in it. Transgenic grain is useful
CC for preparing maltodextrin. A transformed plant (TP) can be used to
CC produce food products having improved taste and to produce fermentable
CC substrates for ethanol and fermented beverages. (M) eliminates the need
CC to mill or physically disrupt the integrity of plant parts prior to
CC recovery of starch-derived products. The present sequence encodes maize
CC 797GL3 granule-bound starch synthase (waxy), which is given in the
CC exemplification of the present invention
XX
SQ Sequence 2223 BP; 435 A; 729 C; 712 G; 347 T; 0 U; 0 Other;

Query Match      3.8%; Score 19; DB 7; Length 2223;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACT 289
Db 2060 GCCGCGGTGCCGACCACT 2042
```

RESULT 21  
ID: AAX63355 standard; cDNA; 2267 BP.  
AC AAX63355;  
XX  
XX 16-JUL-1999 (first entry)  
XX  
XX Granule bound starch synthase encoding cDNA.  
XX  
XX Maize; corn; Zea mays; delta-9 desaturase; GBSS; target; substrate;  
KW granule bound starch synthase; hammerhead ribozyme; hairpin ribozyme;  
KW modulation; gene expression; transgenic plant; cleavage; canola plant;  
KW caffeine synthesis; coffee plant; nicotine production; tobacco;  
KW fruit ripening; flower pigmentation; lignin production; ss.  
XX  
XX Zea mays.  
OS  
XX WO9710328-A2.  
PN  
XX  
XX 20-MAR-1997.  
PD  
XX  
XX 12-JUL-1996; 96WO-US011689.  
PF  
XX  
XX 13-JUL-1995; 95US-0001135P.  
PR  
XX (RIBO-) RIBOZYME PHARM INC.  
PA (DOWC) DOWELANCO.  
XX  
XX Zwick MG, Edington BE, Mcswiggen JA, Merlo PAO, Guo L, Skokut TA;  
PI Young SA, Folkerts O, Merlo DJ;  
PI WPI; 1997-202224/18.  
DR  
XX Ribozyme which modulates plant gene expression - preferably modulates  
PT expression of DELTA-9 desaturase or granule bound starch synthase in  
PT maize or canola.  
PT  
XX  
XX Example 9; Page 31-33; 155pp; English.  
PS  
XX The present invention describes an enzymatic nucleic acid molecule (I)  
CC with RNA cleaving activity, which modulates the expression of a plant  
CC gene. Also described is a gene comprising a cDNA sequence encoding maize  
CC Delta-9 desaturase. (I) can be used to modulate expression of a gene,  
CC preferably Delta-9 desaturase or a granule bound starch synthase (GBSS)  
CC gene, in a plant (preferably a maize or canola plant). (I) can be used to  
CC modulate caffeine synthesis in a coffee plant, nicotine production in a  
CC tobacco plant, fruit ripening processes in an apple, tomato, pear, plum  
CC or peach plant, flower pigmentation in a rose, petunia, chrysanthemum or  
CC marigold plant or lignin production in a tobacco, aspen, poplar or pine  
CC plant  
XX  
XX Sequence 2267 BP; 465 A; 699 C; 733 G; 370 T; 0 U; 0 Other;  
SQ  
Query Match 3.8%; Score 19; DB 2; Length 2267;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 271 GCCGGCGTGGCCGACCACT 289  
DB 1840 GCCGGCGTGGCCGACCACT 1822  
|||||  
RESULT 22  
ACA25719  
ID ACA25719 standard; DNA; 3738 BP.  
XX  
XX ACA25719;  
AC  
XX 19-JUN-2003 (first entry)  
DT  
XX

DE Prokaryotic essential gene #7376.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Burkholderia fungorum.  
OS  
XX WO200277183-A2.  
PN  
XX  
XX 03-OCT-2002.  
PD  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR  
XX 06-SEP-2001; 2001US-00948993.  
PR  
XX 25-OCT-2001; 2001US-0342923P.  
PR  
XX 08-FEB-2002; 2002US-00072851.  
PR  
XX 06-MAR-2002; 2002US-0362699P.  
PR  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
PI  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU21849.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
PT  
XX  
XX Claim 14; SEQ ID NO 13589; 1766pp; English.  
PS  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 3738 BP; 690 A; 1229 C; 1290 G; 529 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 7; Length 3738;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 473 CGGACGGCGCCGCGTCGA 491  
|||||

Db	2387	CGGACGGCGCCGCCGTCTGA	2405
	RESULT 23		
	AAV29752/c		
ID	AAV29752	_standard; DNA; 4800 BP.	
XX	AC	AAV29752;	
XX	AC		
XX	11-SEP-1998	(first entry)	
DT	XX		
XX	Zea mays waxy gene.		
DE	XX		
XX	SER; starch-encapsulating region; fusion vector; glucosyl transferase;		
KW	ss.		
KW	XX		
OS	XX		
XX	Key	Location/Qualifiers	
FH	repeat_region	283..287	
FT		/tag= a	
FT		/note= "direct repeat 1"	
FT	repeat_region	288..292	
FT		/tag= b	
FT		/note= "direct repeat 1"	
FT	repeat_region	293..297	
FT		/tag= c	
FT		/note= "direct repeat 1"	
FT	repeat_region	298..302	
FT		/tag= d	
FT		/note= "direct repeat 1"	
FT	misc_feature	372..385	
FT		/tag= e	
FT		/note= "GC stretch, potential regulatory factor binding"	
FT	misc_feature	442..468	
FT		/tag= f	
FT		/note= "GC stretch, potential regulatory factor binding"	
FT	misc_feature	768..782	
FT		/tag= g	
FT		/note= "GC stretch, potential regulatory factor binding"	
FT	misc_feature	810..822	
FT		/tag= h	
FT		/note= "GC stretch, potential regulatory factor binding"	
FT	misc_feature	821..828	
FT		/tag= i	
FT		/note= "target duplication site (Ac7)"	
FT	CAAT_signal	821..828	
FT		/tag= j	
FT	TATA_signal	867..873	
FT		/tag= k	
FT	misc_feature	887..900	
FT		/tag= l	
FT		/note= "GC stretch, potential regulatory factor"	
FT	exon	901..1080	
FT		/tag= o	
FT		/number= 1	
FT	misc_feature	901	
FT		/tag= m	
FT		/note= "transcriptional start site"	
FT	intron	1081..1219	
FT		/tag= p	
FT		/number= 1	
FT	exon	1220..1553	
FT		/tag= q	
FT		/number= 2	
FT	CDS	1449..4343	
FT		/tag= n	
FT		/product= "glucosyl transferase"	
FT		/note= "contains introns"	
FT	intron	1554..1684	
FT		/tag= r	
FT		/number= 2	
FT	exon	1685..1765	
FT		/tag= s	
FT		/number= 3	
FT	intron	1766..1859	
FT		/tag= t	
FT		/number= 3	
FT	exon	1860..1958	
FT		/tag= u	
FT		/number= 4	
FT	intron	1959..2054	
FT		/tag= v	
FT		/number= 4	
FT	exon	2055..2144	
FT		/tag= w	
FT		/number= 5	
FT	intron	2145..2225	
FT		/tag= x	
FT		/number= 5	
FT	exon	2226..2289	
FT		/tag= y	
FT		/number= 6	
FT	intron	2290..2412	
FT		/tag= z	
FT		/number= 6	
FT	exon	2413..2513	
FT		/tag= aa	
FT		/number= 7	
FT	intron	2514..2650	
FT		/tag= bb	
FT		/number= 7	
FT	exon	2651..2760	
FT		/tag= cc	
FT		/number= 8	
FT	intron	2761..2857	
FT		/tag= dd	
FT		/number= 8	
FT	exon	2858..3101	
FT		/tag= ee	
FT		/number= 9	
FT	intron	3102..3211	
FT		/tag= ff	
FT		/number= 9	
FT	exon	3212..3394	
FT		/tag= gg	
FT		/number= 10	
FT	misc_feature	3358..3489	
FT		/tag= hh	
FT		/note= "target duplication site (Ac9)"	
FT		3395..3489	
FT		/tag= ii	
FT	intron		
FT		/number= 10	
FT	exon	3490..3681	
FT		/tag= jj	
FT		/number= 11	
FT	misc_feature	3570..3572	
FT		/tag= kk	
FT		/note= "target duplication site (Spm 18)"	
FT		3682..3792	
FT	intron		
FT		/tag= ll	
FT		/number= 11	
FT	exon	3793..3879	
FT		/tag= mm	
FT		/number= 12	
FT	intron		

```
FT FT polyA_signal /number= 14
FT 4570..4575
FT /*tag= xx
FT polyA_signal 4593..4598
FT /*tag= ss
FT polyA_signal 4597..4602
FT /*tag= tt
XX XX
XX WO9814601-A1.
XX 09-APR-1998.
XX 30-SEP-1997; 97WO-US017555.
XX 30-SEP-1996; 96US-0026855P.
XX (EXSE-) EXSEED GENETICS LLC.
XX Keeling P, Guan H;
XX WPI; 1998-240100/21.
XX P-FSDB; AAW56484.
XX Hybrid polypeptide comprising starch-encapsulating region and protein -
XX useful for, e.g. producing protein(s) resistant to degradation by stomach
XX acids.
XX Example 2; Page 29-31; 156pp; English.
XX The sequence is that of the waxy gene which codes for glucosyl
XX transferase. It can be used in the production of a hybrid polypeptide
XX comprising a starch-encapsulating region (SER) fused to a payload
XX protein. The hybrid polypeptide can be used to make modified starches
XX comprising the payload protein, selected from, e.g. hormones, growth
XX factors, antibodies, enzymes, dyes, immunoglobulins, etc. The modified
XX starch can also be used to provide grain feeds enriched in amino acids.
XX By encapsulating the payload protein in starch, it is more resistant to
XX degradation by stomach acids
XX SQ Sequence 4800 BP; 935 A; 1414 C; 1446 G; 1005 T; 0 U; 0 Other;
Query Match 3.8%; Score 19; DB 2; Length 4800;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 271 GCCGCGGTGCCGACCACCT 289
DB 4056 GCCGCGGTGCCGACCACCT 4038
RESULT 24
ABX09931/c
ID ABX09931 standard; DNA; 4800 BP.
AC ABX09931;
DT 17-FEB-2003 (first entry)
XX DNA encoding maize granule bound starch synthase (GBSS).
XX Starch; starch synthase; glucan association domain; GLASS; linker domain;
XX LINKR; glucosyl transferase domain; GLYTR; C-terminal end; CTEND;
XX granule bound starch synthase; GBSS; morphology; retrogradation;
XX waterbinding; swelling potential; gene; ds.
XX Zea mays.
XX WO200279410-A2.
XX 10-OCT-2002.
XX 29-MAR-2002; 2002WO-US009574.
XX
```

```
PR 30-MAR-2001; 2001US-0279720P.
XX (BADI ) BASF PLANT SCI GMBH.
XX Commuri P, Keeling PL, Ramirez N, McKean A, Gao Z, Guan H;
XX WPI; 2003-040678/03.
XX New DNA encoding fusion protein consisting of 4 different functional
XX domains selected from glucan association domain, linker domain, glucosyl
XX transferase domain, and C-terminal end, useful for producing modified
XX starches.
XX Claim 32; Page 221-222; 265pp; English.
XX The invention describes an isolated DNA molecule encoding a fusion
XX protein consisting of 4 different functional domains selected from glucan
XX association domain (GLASS), linker domain (LINKR), glucosyl transferase
XX domain (GLYTR), and C-terminal end (CTEND) which are operably linked to
XX one another. The DNA molecule is useful for expressing in plants
XX polypeptides including starch synthase enzymes as fusion proteins with
XX improved affinity to starch and modified catalytic capabilities and to
XX the in vivo and in vitro synthesis of glucan chains of modified lengths
XX as compared to plants producing native starch or starch produced with
XX native starch synthases. Expression of the starch synthase fusion
XX proteins along with granule bound starch synthase (GBSS) will lead to a
XX modified starch having an altered or improved morphology, retrogradation,
XX waterbinding, or swelling potential of the granules, gel strength,
XX adhesiveness, cohesiveness, hardness, elasticity, increased or decreased
XX granule size, degree of branching, crystallinity, degree of cross-
XX linking, and increased or decreased glucan chain lengths. This sequence
XX encodes a starch synthase used in the invention
XX SQ Sequence 4800 BP; 935 A; 1413 C; 1447 G; 1005 T; 0 U; 0 Other;
Query Match 3.8%; Score 19; DB 7; Length 4800;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 271 GCCGCGGTGCCGACCACCT 289
DB 4056 GCCGCGGTGCCGACCACCT 4038
RESULT 25
AAK82712
ID AAK82712 standard; DNA; 6029 BP.
XX AAK82712;
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37524.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
```

PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234957P.  
 PR 25-SEP-2000; 2000US-0234958P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249254P.  
 PR 17-NOV-2000; 2000US-0249255P.  
 PR 17-NOV-2000; 2000US-0249257P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 05-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 XX Disclosure; SEQ ID NO 37524; 3071pp + Sequence Listing; English.  
 XX  
 XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 XX activity, and can be used in gene therapy and vaccine production. (I)  
 XX proteins and polynucleotides may be used in the prevention, diagnosis and  
 XX treatment of diseases associated with inappropriate (I) expression. For  
 XX example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 6029 BP; 1561 A; 1439 C; 1498 G; 1531 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 4; Length 6029;  
Best Local Similarity 100.0%; Pred.No. 29;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 274 GCGGTGCGGACCACTACC 292  
Db 5783 GCGGTGCGGACCACTACC 5801  
RESULT 26  
AAA63350/c  
ID AAA63350 standard; DNA; 21185 BP.  
XX  
AC AAA63350;  
XX  
DT 06-MAR-2001. (first entry)  
XX  
DE Streptomyces globisporus C-1027 gene cluster ORF 25-42.  
XX  
KW Enediyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;  
KW cancer; ds.  
XX  
OS Streptomyces globisporus.  
XX  
FH Key Location/Qualifiers  
FT CDS complement(1..632)  
FT FT  
FT FT  
FT CDS /product= "type II NRPS adenylation enzyme"  
FT 1966..4044  
FT /tag= a  
FT /product= "transmembrane transport protein"  
FT 4188..5192  
FT /tag= c  
FT /product= "O-methyl transferase"  
FT 5249..6505  
FT /tag= d  
FT /product= "P450 hydroxylase"  
FT complement(5539..10361)  
FT /tag= g  
FT /product= "oxidoreductase"  
FT complement(6628..7734)  
FT /tag= e  
FT /product= "oxidoreductase"  
FT 8370..9410  
FT /tag= f  
FT /product= "ORF 31 protein"  
FT 11261..12094  
FT /tag= h  
FT /product= "ORF 33 protein"  
FT 12251..13399  
FT /tag= i  
FT /product= "ORF 34 protein"  
FT 14046..14900  
FT /tag= j  
FT /product= "proline oxidase"  
FT complement(14947..15756)  
FT /tag= k  
FT /product= "ORF 36 protein"  
FT complement(15853..16323)

FT FT  
FT CDS /product= "ORF 37 protein"  
FT complement(16460..18115)  
FT /tag= m  
FT /product= "P450 hydroxylase"  
FT complement(18112..18642)  
FT /tag= n  
FT /product= "ORF 39 protein"  
FT 18960..20039  
FT /tag= o  
FT /product= "ORF 40 protein"  
FT 20065..20919  
FT /tag= p  
FT /product= "ORF 41 protein"  
FT complement(20807..21185)  
FT /tag= q  
FT /product= "ORF 42 protein"  
XX  
PN WO200040596-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US0000446.  
XX  
PR 06-JAN-1999; 99US-0115434P.  
PR 05-JAN-2000; 2000US-00477962.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Shen B, Liu W, Christenson SD, Standage S;  
XX  
XX WPI; 2000-465947/40.  
DR P-PSDB; AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593,  
DR AAB13594, AAB13595, AAB13596, AAB13597, AAB13600, AAB13601,  
DR AAB13602, AAB13603, AAB13607, AAB13606.  
XX  
XX Isolated nucleic acid comprising a nucleic acid encoding any of C-1027  
PT open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for  
PT the production of enediyne C-1027 antitumor antibiotics.  
XX  
PS Claim 1; Page 130-157; 160pp; English.  
XX  
CC The present sequence is the last 21184 bases of the enediyne C-1027 gene  
CC cluster from Streptomyces globisporus. Enediyne C-1027 is an antibiotic,  
CC consisting of an apoprotein and a non-peptidic chromophore, which acts by  
CC damaging DNA. The sequences within the gene cluster, and the proteins  
CC they encode, can be used in the treatment of cancer, along with  
CC antagonists of the protein. Each of the open reading frames is  
CC specifically claimed, excluding ORF 9, which encodes Caga  
XX  
SQ Sequence 21185 BP; 2903 A; 7529 C; 7587 G; 3166 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 3; Length 21185;  
Best Local Similarity 100.0%; Pred.No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 20 CGACCGGCGGCCCAAGG 38  
Db 136 CGACCGGCGGCCCAAGG 118  
RESULT 27  
AAT93095  
ID AAT93095 standard; cDNA; 24379 BP.  
XX  
AC AAT93095;  
XX  
DT 11-MAY-1998 (first entry)  
XX  
DE Streptomyces frenolicin gene cluster.  
XX  
KW Frenolicin; antibiotic; feed additive; anticoccidial; coccidiostatic;  
KW efflux pump; butyrate starter synthase; polyketide synthase; PKS;

KW hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;  
 KN hydroxylase; Streptomyces roseofulvus; ds.  
 OS Streptomyces sp.

XX  
 FH Key  
 FT CDS Location/Qualifiers  
 FT 636..2948  
 FT /\*tag= a  
 FT /product= "80 kDa non-membrane protein"  
 FT /note= "gene A (specifically claimed)"  
 FT 2945..3916  
 FT /\*tag= b  
 FT /product= "membrane protein"  
 FT /note= "gene B (specifically claimed)"  
 FT 4020..4844  
 FT /\*tag= c  
 FT /product= "protein with 6 membrane-spanning domains"  
 FT /note= "gene C (specifically claimed)"  
 FT 4841..6415  
 FT /\*tag= d  
 FT /product= "ATP-binding component of ABC transporter"  
 FT /note= "gene D (specifically claimed)"  
 FT 6533..7183  
 FT /\*tag= e  
 FT /product= "unknown non-membrane protein"  
 FT /note= "gene E (specifically claimed)"  
 FT 7344..8897  
 FT /\*tag= f  
 FT /product= "putative efflux pump"  
 FT /note= "gene F (specifically claimed)"  
 FT 9164..10012  
 FT /\*tag= g  
 FT /product= "transcription activator"  
 FT /note= "gene G (specifically claimed)"  
 FT complement(10105..10621)  
 FT /\*tag= h  
 FT /product= "translationally coupled to gene I"  
 FT /note= "gene H (specifically claimed)"  
 FT complement(10618..11628)  
 FT /\*tag= i  
 FT /product= "homologue of fabH"  
 FT /note= "gene I (specifically claimed)"  
 FT 11809..12066  
 FT /\*tag= j  
 FT /product= "acyl carrier protein"  
 FT /note= "gene J (specifically claimed)"  
 FT complement(12154..13209)  
 FT /\*tag= k  
 FT /product= "acyltransferase"  
 FT /note= "gene K (specifically claimed)"  
 FT 13409..14686  
 FT /\*tag= l  
 FT /product= "PKS ketoacylsynthase subunit"  
 FT /note= "gene L (specifically claimed)"  
 FT 14767..16047  
 FT /\*tag= m  
 FT /product= "PKS chain length factor"  
 FT /note= "gene M (specifically claimed)"  
 FT 16120..16371  
 FT /\*tag= n  
 FT /product= "acyl carrier protein used by the PKS"  
 FT /note= "gene N (specifically claimed)"  
 FT complement(16453..16935)  
 FT /\*tag= o  
 FT /product= "putative hemiketal dehydrase"  
 FT /note= "gene O (specifically claimed)"  
 FT 17088..17903  
 FT /\*tag= p  
 FT /product= "ketoreductase related to actIII"  
 FT /note= "gene P (specifically claimed)"  
 FT 17903..18898  
 FT /\*tag= q  
 FT /product= "cyclase/dehydrase related to act VII"

FT CDS  
 FT /note= "gene Q (specifically claimed)"  
 FT 18895..19839  
 FT /\*tag= r  
 FT /product= "cyclase/dehydrase related to actIV"  
 FT /note= "gene R (specifically claimed)"  
 FT complement(19990..20907)  
 FT /\*tag= s  
 FT /product= "oxidoreductase"  
 FT /note= "gene S (specifically claimed)"  
 FT complement(20904..22094)  
 FT /\*tag= t  
 FT /product= "quinone-forming hydroxylase"  
 FT /note= "gene T (specifically claimed)"  
 FT complement(22505..22179)  
 FT /\*tag= u  
 FT /product= "not specified"  
 FT /note= "gene U"  
 XX  
 PN EP806480-A2.  
 XX  
 PD 12-NOV-1997.  
 PD  
 PF 02-MAY-1997; 97EP-00107329.  
 XX  
 PR 07-MAY-1996; 96US-0016753P.  
 XX  
 PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Reeves CD, Soliday CL;  
 XX  
 DR WPI; 1997-538619/50.  
 DR P-PSDB; AAW34199, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,  
 DR AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,  
 DR AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,  
 DR AAW34219.  
 XX  
 ST Streptomyces frenolicin gene cluster - useful for producing recombinant  
 FT frenolicin antibiotics.  
 FT  
 XX Claim 1; Page 40-60; 66pp; English.  
 XX  
 CC This DNA sequence comprises the Streptomyces frenolicin gene cluster  
 CC containing specifically claimed coding sequences (genes A-U) that  
 CC respectively encode 21 proteins (see AAW34199-219) involved in frenolicin  
 CC synthesis. The genes can be divided into 5 subclusters: (1) genes A, B,  
 CC C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate  
 CC starter synthases; (3) genes L, M and N encode polyketide synthases (PKS)  
 CC ; (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and  
 CC cyclases/dehydrases; and (5) genes S and T encode a keto/enoyl reductase  
 CC and a hydrolase. Also claimed are vectors, host cells (especially a  
 CC Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded  
 CC proteins. Cells transformed using the above sequence can be cultured to  
 CC produce frenolicins or frenolicin precursors. The precursors can be  
 CC converted to frenolicins by chemical or other methods. The frenolicins  
 CC can be oxidised to frenolicin B, an antibiotic used as an anticoccidial  
 CC agent. The frenolicins can be used as animal feed additives  
 XX  
 SQ Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 2; Length 24379;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 477 CGCGCGCCGCGTCGAGGCG 495  
 DB 17279 CGCGCGCCGCGTCGAGGCG 17297  
 RESULT 28  
 AAV25925  
 ID AAV25925 standard; cDNA; 24379 BP.  
 XX  
 AC AAV25925;





FT	CDS	complement(8..658)	FT	FT	/*tag= y	/*product= "sgcB transmembrane efflux protein"
FT	FT	/*tag= a	FT	FT	31280..32590	
FT	CDS	/*product= "ORF -7 protein"	FT	FT	/*tag= z	
FT	FT	complement(930..1478)	FT	FT	/*product= "coenzyme F390 synthetase"	
FT	CDS	/*tag= b	FT	FT	32809..34392	
FT	FT	/*product= "ORF -6 protein"	FT	FT	/*tag= aa	
FT	FT	complement(1649..2713)	FT	FT	/*product= "hydroxylase"	
FT	FT	/*tag= c	FT	FT	complement(34458..35294)	
FT	CDS	/*product= "ORF -5 protein"	FT	FT	/*tag= ab	
FT	FT	complement(2850..3237)	FT	FT	/*product= "ORF 22 protein"	
FT	FT	/*tag= d	FT	FT	/transl_except= (pos:35223..35225,aa:ile)	
FT	FT	/*product= "ORF -4 protein"	FT	FT	/transl_except= (pos:35226..35228,aa:Ala)	
FT	CDS	complement(3442..4971)	FT	FT	complement(35518..35938)	
FT	FT	/*tag= e	FT	FT	/*tag= ac	
FT	CDS	/*product= "ORF -3 protein"	FT	FT	/*product= "ORF 18 protein"	
FT	FT	5982..7479	FT	FT	36165..37490	
FT	FT	/*tag= f	FT	FT	/*tag= ad	
FT	FT	/*product= "glycerol phosphate transporter"	FT	FT	/*product= "coenzyme F390 synthetase"	
FT	CDS	complement(7573..9900)	FT	FT	37559..38938	
FT	FT	/*tag= g	FT	FT	/*tag= ae	
FT	CDS	/*product= "ABC transport/UvrA-like protein"	FT	FT	/*product= "type II NRPS condensation enzyme"	
FT	FT	complement(9982..11349)	FT	FT	38983..39264	
FT	FT	/*tag= h	FT	FT	/*tag= af	
FT	CDS	/*product= "Na+/H+ transporter"	FT	FT	/*product= "type II peptidyl carrier protein"	
FT	FT	complement(11351..12835)	FT	FT	complement(39367..40986)	
FT	FT	/*tag= i	FT	FT	/*tag= ag	
FT	CDS	/*product= "hydroxylase/halogenase"	FT	FT	/*product= "aminomutase"	
FT	FT	13012..14079	FT	FT	complement(41052..42611)	
FT	FT	/*tag= j	FT	FT	/*tag= ah	
FT	FT	/*product= "dNDP-glucose synthase"	FT	FT	/*product= "type II NRPS adenylation enzyme"	
FT	CDS	complement(14212..14643)	FT	FT	43945..46024	
FT	FT	/*tag= k	FT	FT	/*tag= ai	
FT	CDS	/*product= "CagA"	FT	FT	/*product= "transmembrane transport protein"	
FT	FT	complement(14690..15922)	FT	FT	46167..47171	
FT	FT	/*tag= l	FT	FT	/*tag= aj	
FT	FT	/*product= "aminotransferase"	FT	FT	/*product= "O-methyl transferase"	
FT	CDS	complement(15919..16653)	FT	FT	47228..48484	
FT	FT	/*tag= m	FT	FT	/*tag= ak	
FT	CDS	/*product= "N-methyl transferase"	FT	FT	/*product= "P450 hydroxylase"	
FT	FT	complement(16653..17924)	FT	FT	complement(48607..49714)	
FT	FT	/*tag= n	FT	FT	/*tag= al	
FT	FT	/*product= "C-methyl transferase"	FT	FT	/*product= "oxidoreductase"	
FT	CDS	complement(18031..19191)	FT	FT	50350..51390	
FT	FT	/*tag= o	FT	FT	/*tag= am	
FT	CDS	/*product= "epoxide hydrazase"	FT	FT	/*product= "ORF 31 protein"	
FT	FT	complement(19267..19929)	FT	FT	complement(51421..52341)	
FT	FT	/*tag= p	FT	FT	/*tag= an	
FT	CDS	/*product= "anthranilate synthase II"	FT	FT	/*product= "oxidoreductase"	
FT	FT	complement(19926..21407)	FT	FT	53241..54074	
FT	FT	/*tag= q	FT	FT	/*tag= ao	
FT	CDS	/*product= "anthranilate synthase I"	FT	FT	/*product= "ORF 33 protein"	
FT	FT	complement(21424..22878)	FT	FT	54231..55379	
FT	FT	/*tag= r	FT	FT	/*tag= ap	
FT	CDS	/*product= "coenzyme F390 synthetase"	FT	FT	/*product= "ORF 34 protein"	
FT	FT	complement(22875..23546)	FT	FT	56026..56880	
FT	FT	/*tag= s	FT	FT	/*tag= aq	
FT	CDS	/*product= "iron-sulphur flavoprotein"	FT	FT	/*product= "proline oxidase"	
FT	FT	complement(23566..24702)	FT	FT	complement(56927..57736)	
FT	FT	/*tag= t	FT	FT	/*tag= ar	
FT	FT	/*product= "O-acyl transferase"	FT	FT	/*product= "ORF 36 protein"	
FT	CDS	complement(24986..25564)	FT	FT	complement(57833..58304)	
FT	FT	/*tag= u	FT	FT	/*tag= as	
FT	FT	/*product= "epimerase"	FT	FT	/*product= "ORF 37 protein"	
FT	CDS	25815..27170	FT	FT	complement(58440..60095)	
FT	FT	/*tag= v	FT	FT	/*tag= at	
FT	FT	/*product= "monooxygenase"	FT	FT	/*product= "P450 hydroxylase"	
FT	CDS	27214..28593	FT	FT	complement(60092..60621)	
FT	FT	/*tag= w	FT	FT	/*tag= au	
FT	FT	/*product= "glycosyl transferase"	FT	FT	/*product= "ORF 39 protein"	
FT	CDS	28590..29588	FT	FT	60940..62019	
FT	FT	/*tag= x	FT	FT	/*tag= av	
FT	FT	/*product= "dNDP glucose dehydratase"	FT	FT	/*product= "ORF 40 protein"	
FT	CDS	29632..31197				

```

FT CDS      62045..62899
FT          /*tag= aw
FT          /product= "ORF 41 protein"
FT CDS      complement(62787..63164)
FT          /*tag= ax
FT          /product= "ORF 42 protein"
XX WO200040596-A1.
XX 13-JUL-2000.
XX 06-JAN-2000; 2000WO-US000446.
XX 06-JAN-1999; 99US-0115434P.
XX 05-JAN-2000; 2000US-00477962.
XX (REGC ) UNIV CALIFORNIA.
XX Shen B, Liu W, Christenson SD, Standage S;
XX WPI; 2000-465947/40.
XX P-PSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,
XX AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,
XX AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,
XX AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,
XX AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587,
XX AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594,
XX AAB13595, AAB13596, AAB13597, AAB13598, AAB13599, AAB13600, AAB13601, AAB13602,
XX AAB13603, AAB13606, AAB13607.
XX Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for
XX the production of enediyme C-1027 antitumor antibiotics.
XX Claim 1; Page 78-157; 160pp; English.
XX The present sequence is the enediyme C-1027 gene cluster from
XX Streptomyces globisporus. Enediyme C-1027 is an antibiotic, consisting of
XX an apoprotein and a non-peptidic chromophore, which acts by damaging DNA.
XX The sequences within the gene cluster, and the proteins they encode, can
XX
XX Query Match      3.8%; Score 19; DB 3; Length 63164;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY      20 CGACCGGGGGCCCAAGG 38
XX      |||||
XX Db      42115 CGACCGGGGGCCCAAGG 42097
XX
XX RESULT 30
XX ABQ81846/c
XX ID      ABQ81846 standard; DNA; 349980 BP.
XX AC
XX AC      ABQ81846;
XX
XX DT      19-NOV-2002 (first entry)
XX
XX DE      Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1102.
XX
XX KW      Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition; gene; db.
XX
XX OS      Bifidobacterium longum.
XX
XX OS      Synthetic.
XX
XX PN      EP1227152-A1.
XX
XX 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX
XX 30-JAN-2001; 2001EP-00102050.
XX (NEST ) SOC PROD NESTLE SA.
XX WPI; 2002-668397/72.
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.
XX Disclosure; SEQ ID NO 1102; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
XX sequence of a Bifidobacterium genome selected from the nucleotide
XX sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
XX least 90% identity or which hybridises with the sequences given in
XX ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
XX fusion protein, comprising a sequence selected from 1097 sequences given
XX in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
XX heterologous polypeptide. (I) has antidiarrheic and antibacterial
XX activities, and can be used as an inhibitor of Salmonella. (I) (which is
XX a probe) is useful for the detection and/or identification of
XX Bifidobacterium longum in a biological sample. A carrier containing the
XX lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM 1-2618) can be
XX used for preventing and/or treating diarrhoea brought about by pathogenic
XX bacteria and/or rotavirus. The carrier is a food composition selected
XX from milk, yogurt, curd, cheese, fermented milks, milk based fermented
XX products, ice-creams, fermented cereal based products, milk based
XX powders, infant formula, pet food or a pharmaceutical composition
XX selected from tablets, liquid bacterial suspensions, dried oral
XX supplement, wet oral supplement, dry tube feeding or wet tube feeding.
XX (I) is useful in DNA arrays or chips to carry out analysis of the
XX expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
XX Bifidobacterium related nucleotide sequences given in the Sequence
XX listing from the present invention but not mentioned further within the
XX specification. N.B. The sequence data for this patent is not represented
XX in the printed specification but is based on sequence information
XX supplied by the European Patent Office
XX
XX Sequence 349980 BP; 69200 A; 103414 C; 105147 G; 72219 T; 0 U; 0 Other;
XX
XX Query Match      3.8%; Score 19; DB 6; Length 349980;
XX Best Local Similarity 100.0%; Pred. No. 18;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY      219 CTCGCGCGACGACGTGATG 237
XX      |||||
XX Db      247972 CTCGCGCGACGACGTGATG 247954
XX
XX RESULT 31
XX AAI29319
XX ID      AAI29319 standard; cDNA; 175 BP.
XX
XX AC      AAI29319;
XX
XX DT      12-OCT-2001 (first entry)
XX
XX DE      Colon tumour related determined cDNA sequence for clone R0095:F10.
XX
XX KW      Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
XX gene therapy; vaccine; colonic cancer; ss.
XX
XX OS      Homo sapiens.
XX
XX OS      WO200149716-A2.
XX
XX PN      12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US035596.
XX
XX 30-DEC-1999; 99US-00476296.
XX

```

```
PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.
PR 28-AUG-2000; 2000US-00649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX WPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 2; Page 376; 472pp; English.
XX
XX The present invention describes colon tumor associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumor associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be used
XX to treat disorders associated with decreased expression by rectifying
XX mutations or deletions in a patient's genome that affect the activity of
XX TCAPs by expressing inactive proteins or to supplement the patients own
XX production of them. Additionally, (II) may be used to produce the TCAP
XX proteins, by inserting the nucleic acids into a host cell culturing the
XX cell to express the protein. (II) and its complementary sequences may
XX also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
XX and hybridisation assays to detect and quantitate the presence of similar
XX nucleic acids in samples, and therefore which patients may be in need of
XX restorative therapy. (I) may also be used as antigens in the production
XX of antibodies against TCAPs and in assays to identify modulators of TCAP
XX expression and activity. Anti-(I) antibodies and antagonists may also be
XX used to down regulate TCAP expression and activity. The anti-(I)
XX antibodies may also be used as diagnostic agents for detecting the
XX presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
XX (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
XX nucleotide and amino acid sequences given in the exemplification of the
XX present invention
XX
XX SQ Sequence 175 BP; 30 A; 70 C; 56 G; 19 T; 0 U; 0 Other;
XX
XX Query Match 3.6%; Score 18; DB 4; Length 175;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 297 CCTGGTGGCGGCCCG 314
XX |||||||
XX Db 36 CCTGGTGGCGGCCCG 53
XX
XX RESULT 32
XX ABN18020
XX ID ABN18020 standard; cDNA; 313 BP.
XX
XX XX
XX AC ABN18020;
XX
XX XX
XX DT 24-JUN-2002 (first entry)
XX
XX XX
XX DE Human ORFX polynucleotide sequence SEQ ID NO:4517.
XX
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis; gene; ss.
XX
XX OS Homo sapiens.
XX
XX XX
XX PN WO200192523-A2.
XX
XX XX
XX PD 06-DEC-2001.
XX
XX XX
XX PF 29-MAY-2001; 2001WO-US010836.
XX
XX XX
XX PR 30-MAY-2000; 2000US-0206132P.
XX 29-AUG-2000; 2000US-0228716P.
XX
XX XX
```

```
PR 10-JAN-2000; 2000US-00480321.
PR 15-FEB-2000; 2000US-00504629.
PR 06-MAR-2000; 2000US-00519444.
PR 19-MAY-2000; 2000US-00575251.
PR 29-JUN-2000; 2000US-00609448.
PR 28-AUG-2000; 2000US-00649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX WPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 2; Page 376; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be used
XX to treat disorders associated with decreased expression by rectifying
XX mutations or deletions in a patient's genome that affect the activity of
XX TCAPs by expressing inactive proteins or to supplement the patients own
XX production of them. Additionally, (II) may be used to produce the TCAP
XX proteins, by inserting the nucleic acids into a host cell culturing the
XX cell to express the protein. (II) and its complementary sequences may
XX also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
XX and hybridisation assays to detect and quantitate the presence of similar
XX nucleic acids in samples, and therefore which patients may be in need of
XX restorative therapy. (I) may also be used as antigens in the production
XX of antibodies against TCAPs and in assays to identify modulators of TCAP
XX expression and activity. Anti-(I) antibodies and antagonists may also be
XX used to down regulate TCAP expression and activity. The anti-(I)
XX antibodies may also be used as diagnostic agents for detecting the
XX presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
XX (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
XX nucleotide and amino acid sequences given in the exemplification of the
XX present invention
XX
XX SQ Sequence 175 BP; 30 A; 70 C; 56 G; 19 T; 0 U; 0 Other;
XX
XX Query Match 3.6%; Score 18; DB 4; Length 175;
XX Best Local Similarity 100.0%; Pred. No. 1.3e+02;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 297 CCTGGTGGCGGCCCG 314
XX |||||||
XX Db 36 CCTGGTGGCGGCCCG 53
XX
XX RESULT 32
XX ABZ33505
XX ID ABZ33505 standard; cDNA; 175 BP.
XX
XX XX
XX AC ABZ33505;
XX
XX XX
XX DT 30-JAN-2003 (first entry)
XX
XX XX
XX DE Human colon tumour cDNA for clone R0095.F10 SEQ ID NO:873.
XX
XX XX
XX KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
XX
XX XX
XX OS Homo sapiens.
XX
XX XX
XX PN WO200283070-A2.
XX 24-OCT-2002.
XX
XX XX
```

PA (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach MD;  
XX WPI; 2002-106308/14.  
DR P-PSDB; ABP02268.  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.  
XX Disclosure; SEQ ID NO 4517; 1037pp; English.  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative diseases, diabetes mellitus, systemic  
CC transplantation, cardiovascular diseases, diabetes mellitus, cholesterol  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 313 BP; 61 A; 107 C; 87 G; 58 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 6; Length 313;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 211 GACGCGAGCTCGCGGAC 228  
Db ||||||||||||||||  
223 GACGCGAGCTCGCGGAC 240  
  
RESULT 34  
ACH19833  
ID ACH19833 standard; cDNA; 418 BP.  
XX AC ACH19833;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
DE Human adult lung cDNA #836.  
XX  
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2003073623-A1.  
XX  
PD 17-APR-2003.  
XX  
PF 30-JUL-2001; 2001US-00918995.  
XX  
PR 30-JUL-2001; 2001US-00918995.  
XX  
PA (DRMA/) DRMANAC R T.  
  
(LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
XX mapping, in the recombinant production of protein, or in generating  
XX antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 7045; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
XX determined by the technique of SBH (sequencing by hybridisation). Also  
XX included is a purified polypeptide comprising a sequence corresponding to  
XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
XX are useful in diagnostics as expressed sequence tags (EST) for  
XX identifying expressed genes or for physical mapping of the human genome,  
XX in forensics, in assessing biodiversity, or in identifying mutations  
XX responsible for genetic disorders and other traits. The nucleotide  
XX sequences are also useful as hybridisation probes, as oligomers for PCR,  
XX for chromosome and gene mapping, in the recombinant production of  
XX protein, or in generating antisense DNA or RNA. The purified polypeptide  
XX is useful for generating antibodies specific for it. The present sequence  
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
SQ Sequence 418 BP; 89 A; 128 C; 133 G; 68 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 8; Length 418;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 350 GCCTGCGCGGCGCGCG 367  
Db ||||||||||||||||  
309 GCCTGCGCGGCGCGCG 326  
  
RESULT 35  
AAH83034  
ID AAH83034 standard; cDNA; 457 BP.  
XX AC AAH83034;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:658.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 16-JAN-2001; 2001WO-US001575.  
XX  
PR 14-JAN-2000; 2000US-0176722P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Algate PA;  
XX  
XX WPI; 2001-425866/45.  
XX  
DR

XX Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer.  
PT  
XX  
PS Claim 5; Page 185; 339pp; English.  
XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit the  
CC development of cancer, particularly ovarian cancer. They can also be used  
CC to diagnose the onset and progression of cancer  
XX  
SQ Sequence 457 BP; 127 A; 83 C; 87 G; 157 T; 0 U; 3 Other;  
  
Query Match 3.6%; Score 18; DB 5; Length 457;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 23 CCGGGCGGCCCAAGGGCG 40  
|||||  
Db 367 CCGGGCGGCCCAAGGGCG 384  
  
RESULT 36  
AAV29373  
ID AAV29373 standard; cDNA; 498 BP.  
XX  
AC AAV29373;  
XX  
DT 03-AUG-1998 (first entry)  
XX  
DE Human small nuclear ribo-nucleoprotein Sm protein (HSMNP) encoding cDNA.  
XX  
KW Small nuclear ribo-nucleoprotein; snRNP; Sm protein; HSMNP; human;  
KW diagnosis; prevention; treatment; systemic lupus erythematosus; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
PT CDS 28..339  
FT /\*tag= a  
FT /product= "HSMNP protein"  
XX  
XX W09813380-A1.  
XX  
PD 02-APR-1998.  
XX  
XX 26-SEP-1997; 97WO-US017362.  
XX  
XX 27-SEP-1996; 96US-00722349.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hallman JL, Bandman O, Zweiger GB;  
XX  
XX WPI; 1998-230633/20.  
XX  
XX P-PSDB; AAW56788.  
XX  
PT New isolated human small nuclear ribo-nucleoprotein Sm proteins - is used  
PT to develop products for diagnosis, study, prevention and treatment of  
PT disease, particularly systemic lupus erythematosus.  
XX  
XX Claim 3; Fig 1A-B; 59pp; English.  
XX  
XX This cDNA encodes a human small nuclear ribonucleoprotein (snRNP) Sm  
XX protein (HSMNP) A. A recombinant host cell containing a vector comprising  
XX polynucleotide sequences encoding HSMNP or HSMNPB can be used to produce  
XX these proteins. The HSMNP and HSMNPB and its products can be used in the  
XX diagnosis, prevention and treatment of systemic lupus erythematosus. In  
XX particular agents which modulate the activity of HSMNP would be useful  
XX therapeutically for decreasing the harmful effects of the anti-Sm immune

CC response in systemic lupus erythematosus  
XX  
SQ Sequence 498 BP; 112 A; 143 C; 151 G; 92 T; 0 U; 0 Other;  
  
Query Match 3.6%; Score 18; DB 2; Length 498;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 350 GCCTGGCGCGGGCGGCGG 367  
|||||  
Db 339 GCCTGGCGCGGGCGGCGG 356  
  
RESULT 37  
AAS75981  
ID AAS75981 standard; cDNA; 537 BP.  
XX  
AC AAS75981;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #11785.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
DN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX P-PSDB; ABG11794.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
XX  
XX Claim 1; SEQ ID NO 11785; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
XX coding sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

```
SQ Sequence 537 BP; 80 A; 198 C; 189 G; 68 T; 0 U; 2 Other;
Query Match 3.6%; Score 18; DB 5; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGGCGCCCG 314
Db 327 CCTGGTGGGCGCCCG 344

RESULT 38
AAL01672
ID AAL01672 standard; cDNA; 622 BP.
XX
XX AC
XX AAL01672;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SEQ ID NO: 1673.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
```

PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0243218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-465570/50.  
DR P-PSDB; AAM95702.  
DR  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
PT  
XX Claim 1; SEQ ID NO 1673; 1297pp + Sequence Listing; English.  
PS  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention  
XX  
XX Sequence 622 BP; 181 A; 166 C; 169 G; 104 T; 0 U; 2 Other;  
SQ  
Query Match 3.6%; Score 18; DB 4; Length 622;  
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 350 GCCTGGCGGGGGCGCG 367  
Db 409 GCCTGGCGGGGGCGCG 426  
RESULT 39  
ABA07591  
ID ABA07591 standard; cDNA; 622 BP.  
XX  
AC ABA07591;  
XX  
XX 11-JAN-2002 (first entry)  
XX  
XX Human ovarian and breast cancer associated polynucleotide SEQ ID NO 148.  
DE  
XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200155325-A2.  
XX  
XX 02-AUG-2001.  
PD

XX 17-JAN-2001; 2001WO-US001345.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.



29-SEP-2000; 2000US-0236369P.  
29-SEP-2000; 2000US-0236370P.  
02-OCT-2000; 2000US-0236802P.  
02-OCT-2000; 2000US-0237037P.  
02-OCT-2000; 2000US-0237038P.  
02-OCT-2000; 2000US-0237039P.  
02-OCT-2000; 2000US-0237040P.  
13-OCT-2000; 2000US-0239935P.  
13-OCT-2000; 2000US-0239937P.  
20-OCT-2000; 2000US-0240960P.  
20-OCT-2000; 2000US-0241221P.  
20-OCT-2000; 2000US-0241785P.  
20-OCT-2000; 2000US-0241786P.  
20-OCT-2000; 2000US-0241787P.  
20-OCT-2000; 2000US-0241808P.  
20-OCT-2000; 2000US-0241809P.  
20-OCT-2000; 2000US-0241826P.  
01-NOV-2000; 2000US-0244617P.  
08-NOV-2000; 2000US-0246474P.  
08-NOV-2000; 2000US-0246475P.  
08-NOV-2000; 2000US-0246476P.  
08-NOV-2000; 2000US-0246477P.  
08-NOV-2000; 2000US-0246478P.  
08-NOV-2000; 2000US-0246523P.  
08-NOV-2000; 2000US-0246524P.  
08-NOV-2000; 2000US-0246525P.  
08-NOV-2000; 2000US-0246526P.  
08-NOV-2000; 2000US-0246527P.  
08-NOV-2000; 2000US-0246528P.  
08-NOV-2000; 2000US-0246532P.  
08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249244P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249264P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
06-DEC-2000; 2000US-0256713P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-488786/53.  
P-PSDB; AB010880.

XX New isolated ovarian and/or breast cancer related nucleic acids and  
PT polypeptides, useful for diagnosing, treating and/or preventing human  
PT diseases and disorders, particularly ovarian and/or breast cancer.  
XX Claim 1; SEQ ID NO 148; 577pp + Sequence Listing; English.  
PS  
XX The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (AB010743-AB010980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
XX  
SQ Sequence 622 BP; 181 A; 166 C; 169 G; 104 T; 0 U; 2 Other;  
Query Match 3.6%; Score 18; DB 4; Length 622;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 350 GCCTGGCGGGCGCGCG 367  
Db 409 GCCTGGCGGGCGCGCG 426  
|||||||  
RESULT 40  
AAS81267  
ID AAS81267 standard; cDNA; 633 BP.  
XX  
AC AAS81267;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17071.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; AB017080.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 17071; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridization probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activities. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 633 BP; 151 A; 189 C; 181 G; 112 T; 0 U; 0 Other;  
 Query Match 3.6%; Score 18; DB 5; Length 633;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGGGGGGGCGG 367  
 |||||  
 Db 454 GCCTGGCGGGGGGGCGG 471

RESULT 41  
 ACC83519  
 ID ACC83519 standard; cDNA; 819 BP.  
 XX  
 AC ACC83519;  
 XX  
 DT 08-SEP-2003 (first entry)  
 XX  
 DE Human exosome component hMTR3 coding sequence.  
 XX  
 KW Human; exosome; hMTR3; mRNA turnover; antiinflammatory; cytostatic; gene;  
 KW ss.  
 XX Homo sapiens.  
 OS  
 XX WO2003044166-A2.  
 FN  
 XX 30-MAY-2003.  
 PD  
 XX  
 PF 14-NOV-2002; 2002WO-US036665.  
 XX  
 PR 15-NOV-2001; 2001US-0334712P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chen C, Karin M;  
 XX  
 DR WPI; 2003-457602/43.  
 DR P-PSDB; ABR42950.  
 XX  
 PT New purified mammalian exosome, useful for treating immune diseases e.g.  
 PT diabetes, cirrhosis, scleroderma, lupus, arthritis, or multiple  
 PT sclerosis, inflammatory diseases, cancer, or hormone deficiency disease  
 PT e.g. osteoporosis.  
 XX  
 XX Disclosure; Fig 3D; 127pp; English.  
 PS  
 XX The present sequence is the coding sequence of the human exosome  
 CC component, hMTR3. Peptide fragments (see ABR42935-38) of hMTR3 were

CC identified in a protein isolated from the human exosome. The human  
 CC exosome promotes degradation of AU-rich element (ARE)-containing RNAs.  
 CC ARE recognition by the exosome requires ARE-binding proteins that  
 CC interact with the exosome and recruit it to unstable RNAs, thereby  
 CC promoting their rapid degradation. A claimed method for altering the  
 CC level of an ARE-containing RNA involves contacting the RNA with an ARE-  
 CC binding protein, a mammalian (human) exosome and an agent that alters the  
 CC level of binding of the exosome to the ARE-binding protein. Levels of  
 CC mRNA are decreased or increased by impacting the level of degradation of  
 CC the mRNA. Reducing the level of a cytokine, proto-oncogene or growth  
 CC factor RNA can be used to treat an immune disease, inflammatory disease  
 CC or cancer. Increasing the level of a target RNA can be performed in a  
 CC subject capable of developing a hormone deficiency

XX SQ Sequence 819 BP; 87 A; 311 C; 330 G; 91 T; 0 U; 0 Other;  
 Query Match 3.6%; Score 18; DB 8; Length 819;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGGCGCGCG 314  
 |||||  
 Db 774 CCTGGTGGCGGGCGCGCG 791

RESULT 42  
 ACA44265/c  
 ID ACA44265 standard; DNA; 909 BP.  
 XX  
 AC ACA44265;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #25922.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Pseudomonas putida.  
 XX  
 FN WO200277183-A2.  
 EN  
 XX 03-OCT-2002.  
 PD  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU40395.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 32135; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 909 BP; 188 A; 275 C; 264 G; 182 T; 0 U; 0 Other;  
Query Match 3.6%; Score 18; DB 7; Length 909;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 CGAGCGGAGAGCTCGAC 263  
|||||  
DB 294 CGAGCGGAGAGCTCGAC 277  
|||||

RESULT 43  
AAH14082  
ID AAH14082 standard; cDNA; 1699 BP.  
XX AC AAH14082;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA sequence SEQ ID NO:11239.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX FN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-00116126.  
XX PR 29-JUL-1999; 99JP-00248036.  
XX PR 27-AUG-1999; 99JP-00300253.  
XX PR 11-JAN-2000; 2000JP-00118776.  
XX PR 02-MAY-2000; 2000JP-001183767.  
XX PR 09-JUN-2000; 2000JP-00241899.  
XX PA (HELI-) HELIX RES INST.  
XX OTa T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX PR Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.

XX Claim 8; SEQ ID NO 11239; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesising 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX SQ Sequence 1699 BP; 338 A; 556 C; 440 G; 365 T; 0 U; 0 Other;  
Query Match 3.6%; Score 18; DB 4; Length 1699;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CCACCTACCACCACTGG 301  
|||||  
DB 955 CCACCTACCACCACTGG 972  
|||||

RESULT 44  
ABL90390  
ID ABL90390 standard; cDNA; 1703 BP.  
XX AC ABL90390;  
XX DT 24-MAY-2002 (first entry)  
XX DE Human polynucleotide SEQ ID NO 952.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;  
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; human; secreted protein; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200190304-A2.  
XX PD 29-NOV-2001.  
XX PF 18-MAY-2001; 2001WO-US016450.  
XX PR 19-MAY-2000; 2000US-0205515P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Birse CE, Rosen CA;  
XX WPI; 2002-122018/16.  
XX P-PSDB; ABB89981.  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
XX prevention of neural, immune system, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative

disorders.

Claim 4; SEQ ID NO 952; 2081pp + Sequence Listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

Sequence 1703 BP; 324 A; 569 C; 436 G; 364 T; 0 U; 10 Other;

Query Match 3.6%; Score 18; DB 6; Length 1703;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CCACCTACCACCACTGG 301  
|||  
Db 1063 CCACCTACCACCACTGG 1080

RESULT 45  
AAZ24887

ID RAZ24887 standard; DNA; 1715 BP.

AC AAZ24887;

XX

XX

DT 02-DEC-1999 (first entry)

XX

DE Human secreted protein gene 77 clone HPXKT05.

XX

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX

OS Homo sapiens.

XX

XX WO9947540-A1.

XX

XX 23-SEP-1999.

XX

XX 18-MAR-1999; 99WO-US005804.

XX

PR 19-MAR-1998; 98US-0078563P.  
PR 19-MAR-1998; 98US-0078566P.  
PR 19-MAR-1998; 98US-0078573P.  
PR 19-MAR-1998; 98US-0078574P.  
PR 19-MAR-1998; 98US-0078576P.  
PR 19-MAR-1998; 98US-0078577P.  
PR 19-MAR-1998; 98US-0078578P.  
PR 19-MAR-1998; 98US-0078579P.  
PR 19-MAR-1998; 98US-0078581P.  
PR 01-APR-1998; 98US-0080312P.  
PR 01-APR-1998; 98US-0080313P.  
PR 01-APR-1998; 98US-0080314P.

XX

(HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Peng P, Soppet DR;  
XX Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
PI Shi Y, Moore PA;  
DR WPI: 1999-562050/47.  
DR P-PSDB; AAY41384.  
XX  
XX New isolated human genes, useful for diagnosis and treatment of e.g.  
PT cancers, neurological disorders, immune diseases, inflammation or blood  
disorders.  
PS Claim 1; Page 344-345; 484pp; English.  
XX  
XX This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. AAZ24802) for increasing the stability of the fused protein as  
CC compared to the human protein only. The invention relates to 35 novel  
CC genes and their fragments (nucleic acid sequences: AAZ24811-Z24907; amino  
CC acid sequences AAY41308-Y41404) which are useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
CC pathological conditions can be diagnosed by determining the amount of the  
CC new polypeptides in a sample or by determining the presence of mutations  
CC in the new polynucleotides. Specific uses are described for each of the  
CC 95 polynucleotides, based on which tissues they are most highly expressed  
CC in (see AAZ24811 for described uses)  
XX  
SQ Sequence 1715 BP; 356 A; 557 C; 443 G; 359 T; 0 U; 0 Other;

Query Match 3.6%; Score 18; DB 2; Length 1715;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 CCACCTACCACCACTGG 301  
|||  
Db 950 CCACCTACCACCACTGG 967

Search completed: June 16, 2004, 17:43:50  
Job time : 252 secs



synthase gene cluster from Amycolatopsis mediterranei  
 Unpublished  
 REFERENCE 2 (bases 1 to 53784)  
 AUTHORS Schupp, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel, SWITZERLAND

FEATURES  
 source 1..53784  
 /organism="Amycolatopsis mediterranei"  
 /mol\_type="genomic DNA"  
 /strain="LBG A3136"  
 /db\_xref="taxon:33910"  
 1336..15543  
 /function="CoA ligase, acyltransferase, ketoreductase"  
 b-ketocyl-ACP-synthase, acyltransferase, ketoreductase"  
 /note="ORF1"  
 PKS modules 1-3"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rifamycin polyketide synthase, type 1"  
 /protein\_id="CA11035.1"  
 /db\_xref="GI:2764761"  
 /db\_xref="GOA:O54666"  
 /db\_xref="SPTREMBL:O524666"  
 /translation="MRTDLIKPLHVLLENATRFAGKPAFADHRTVTYTGDLKARTLR  
 LAGHLAGLVGRHGRVRAICLGNRVSTVESYFAILRACGAVGVLNPGSGATAELEHPLTD  
 SGATVVVDAQAARLARLAPHEVLLVTGDDVPGEAGHSYDELAISEPAEPAADLELDE  
 PAMFYSITGTPKGVVSTORNCIMSVASCYVFPFGSDQDRDQWLPFLPFLSHLHIA  
 CWSVTSVARSADGSSADDVNRLEAESSTFLAGVPTTYHHLVRAARQGFSAF  
 LRGLAGAVLGLRSEFETFCVPLIDAYGSETTCGATIMNPPDCAWGVEGSGCLAV  
 PGVDRVVDPTGLDPAGEGEVWSPGNMGLYHNSPEATAAMDDGWFRTGDLIAR  
 RDAGYPTICGRLEKELIRGCANTHPCGEVRLVTDGVADAAGVGVPHDTLCEPVA  
 VYIPGTPGPAAHLEKCRQLSNKYKVPDRILEVAHLPRTASGKIRKGLLTDPAQR  
 YATEHEQSGRVAALAKRLSGDERAQCELEDIVRTQAADVQLDQVDPGR  
 FRDGTSLAIVELNRKLTHETGLWLPASVDFPHTPAALAAARVRAELLGITQAVAP  
 VVAADPEPTAIVGMACRLPGVASPELDLWLRVAERVDVAEFPDGRGMDLSLIDPD  
 RERAGTYSVGGGFLHDAGFDAGFGGISPREAVAMPDQORLLLETSWEALENAGVDP  
 IALKGTDTPGVFGMGCGSGGAVAPLEGFTVTGVASVSGRVSVLGLEGPATV  
 DTACSSSLVAMHIAAQAALRGSCSMALAGVTWATPGSFVFSRQALAPDRCXAF  
 AAADGTGSGEGVVLRSLVAREGRHILAVIRGSAVNQDASNGSLTAPNLSQ  
 RVTRRALAGLAPSDVDVVEAGTGTGLDPTAEQALLATYGOERQPLWGLSKSN  
 IGHAAQAGVAGIKMQLRLEHPTLPTLHDKPTLEVDSAGAILLELTARAWPRNG  
 RPRRAGSVFGVSGTNAHLILEAPAEPEVAAPELPPVPLVWSARSTESLGOAERLA  
 SLLEGVSLTEVAGALVSRRAVLDERAVVAGSFEAEVTLGLRALTAGSGTGGKVVV  
 FPGGTQWAGMRELLAESPVFAERIAECGAALAPWIDNSLVDVLRGEGDLGRVVLQ  
 PACFVWGLAAVWESGVDPDAVVGSHQGEIAAACVSGALSLEDAAKVALRSQAIA  
 AELSGGMSALGEDDVSRILVDGVEVAAVNGPSVVIAGDAHALDAILLSGEG  
 IRVRRVADYASHTHRVEDIRDTLAETLAGI SAQAPVFPYTSVSEWRDAGVLDGG  
 YRRLNKNQVFGAAAVALLEQGHVFEVSAHPVTQPLSELGTDAIGTLRREDGGL  
 RLLASMGELFVRGIDVDMTAMVPAACGWDLPTVAFRHHVWLEPAPASAGDPLICT  
 VVSTPGSRLTAVAQMSERAPQAPWDCGLVFNALVEAAILGLDAGTTPVVELVDP  
 VVLPGRSREVQLIIGBPGCRPPIEFVREADPEWIRHAGTLAPAAAVPEPAAA  
 GDATVTVAGLRDADRYGIHPLDAAVTVVGDLLPSVMTGVSLLASGATVATVP  
 TATGLRITDPAGQPLTVESVGPFFVAEQGTDDALFRVDMPELPTAEATPLPYE  
 ATGAETLSALQAWLADPAETRLAVVTGCTEPCGAAILWGLVRSQAQSEHPICRIVLADL  
 DDPAVLTAHVSGEPQVVRNGVASVPRLTRVTPQDARPLDPEGTVLITGGTGLCA  
 LTHAHLTAHVSGEHLVLSRGAPELOBELTALGASVTAACDVAQLEAVLRAI  
 PAEHLTAHLTAGLDDGVVTEPDRLATVRRPKVDARLLDREADLAEVILF  
 SSAAGLVNPGQAGAAANAEALDALARNSLDLPAYSIAWGYWATVAGSGVTEHPLDGL  
 LRNRQICMSGLPADEGMALIDIAIATGGTILVAKFDPVAAALRAKAGGVPVPLLRGL  
 APIPLRAAKTASLTIRLAGLAETEQAAILDLVREHAEVLHSGASVNRPAEITFKO  
 AGPDSLTAVELNRLLAATGILTSPMIEDYKPKPALADHLRAKFGSAANRPAEITC  
 AAABEPTAIVMACRPFPGVHSPEDLRLVADGADATVTEPADRCNWDTRLIHEDDPH  
 EGTTYVRHGFALDDAAGFDAAFGISPEALAMPDQORLLLETLSWELFEAAIDPTTL  
 AGQDIPYHGFVNSHDYSMRHAAGVEGFRITGGASVLSGRVAFVGEVGPATVIT  
 ACSSGLVAMVALQALRGECMSMALAGVMMVGTVEFVFSRQGLADPGRCAKFAD  
 GADGTGSEGVGLLVERLSEAQRHQHVLAVVRGSAVNSDGSNGHTAPNPSQOVR  
 IRKALAAGTSDVDAVEAGTGTGLDPLTEAALLATYQNEETPLWLSGVSKNLG  
 HTQAAGVAGVIKMMAMRHGVLPRTHLHVRPSSYVDWSAGAVELLTEARDVNSGHP  
 RQAGSVSGTGNNAHVREVAAPITTPQEPPEAFILPVILVSAAGVALLAEHAGLQGRILA  
 AFLGDRDTRVVPDAAYALATTRAQLDHRVVLASDRACLADLAFAFGSGVVTGTFVDG

CDS

KLAVFTGQGSOWAGMRELAETFFVFRDAFEAAACEAVDTHLRERPLREVVDFDSALL  
 DOTMYTCGALFVETALFRLFSBQVRPGLLAGHSIGELAAAHVSGVLDLADACELVA  
 ARGRLMOALPAGGAMVAVQATEDEVALPDGTVCVAANGPDSVVLSTGTAALVAD  
 ELAGRKRTERLAVSHAFSPLEMLDDPRAVAERLTYRAGSLPVYSTLTGELAALD  
 SPYVQGVQVNAVRFSDAVTAUGAQGASTFELGPGGALAAALMALTGGFQSCVATL  
 RKGAEVDPVDLTALAEHLVGVGVDWTTVLDEPATAVGTVLPTVFOHOFKDFWSDVET  
 AAVSVTPPFAEPTVDRPVQDLVRESAAVILGHRDAGSFDLDRSKDFHGFDSLSAV  
 KNLRLRDFTFGVELPSTLI PDYPNPAVLADHLRAELIGERPAAPVPAVTPHAGTSVTRG  
 VGMSTRLPGSGDSPEELKLVARGDAVSGFPVDRGMDLGLVHPDPAHAGTSVTRG  
 GFTHDAQFDPAGLPGISPREALAMPQORLLLETSMELERAGVDPPLSARGSDGVFT  
 GIVHDDYTKLRVEDPVQGTWTGTASVSARGVAIVFGFEGPAVTVDTACSSSLVA  
 MHIAAQALRGECMSMALAGATVMSADAFLEFSRQRLSGADRCAYASAGDGTGWA  
 EGVGVVLERLSVARERGRHVLAVLRGSVNQDASNGLTAPNPSQOVRIRGALASA  
 GLAPSDVVEVGEHGTGALDPIEVQALLATYGOERQPLWGLSKNLGHTQAAAGV  
 VGVIMAMRHGMPATLHVDRSTQVDSAGAEIVLTAREWPRTRGRPRRAGVSF  
 GASGTNAHLIIESGPAEEADEVESVPLVVSARSAGSLAGQAALAAVLENSLAG  
 VAGALVSGRATINERAVVITAGSRDAQGLQALARGENAPGVVTGTAGKGVVWVFP  
 GQGSQWMMGRDILDSVFAARIKECAALEQWTDWLSLLDRLGADLDRVDVQF  
 ASFAMVGLAAVMTSLGTPDAVLHSHQGEIAAACVSGALSLLDDAAKVALRSQAIA  
 ELARGGMSALVSEEDAVARLTPWANRVEVAAVNSPSSVVIAGDAQALDEALELAG  
 DGVRRVAVDYASHTHRVEAIAETLAKTLAGIDARVPAIPFPYSTLTGTEQAVVDA  
 GMYRNLROQVRGSPVADLAGHGVTFVEISAPHLVQPLSEISDDAVTVTRRDD  
 GGLRLLSAELHYVRGVAVDTRAAPAAGVNDLPTVAPRRRHFHLHEATRAEAGM  
 DGEFTWAEQSDVDSLAELLEPEQREGALSTVVPVLAQRDRRRRSTAEKLRQVT  
 WQFLERAGVPGGRMLAVVPAAGTTTALLKELTQGGDLTVLRETEASRAQLAQLRN  
 VLAHDITGLVSLALDGPADAEEITASTLALVQALGDTTISAPLCLTSGAVNIGI  
 QDAVTAQAQAVMLGRVAPERLDRWGLVDLPAAIDARTAOALLGVLNCAAGEDOL  
 AVRSGVYRRLVRKVPESAITSRWRPGTVLVTGGAEGLGRHASVLAQSGAERLLV  
 TGTGDVDELTAELAEFGTTVEFCADTRDAIAQLVADSEVTVAVVHAADIAQTSVDDT  
 GVADLVEFAKVTVAWLDQDFETPLDAFVVFSSITAGIWWGGGQFPAGANAVLDA  
 LFWMRARGALKTSIAMGALDQIGIMDEAALQRRRGTVPMAPLPVAWQAVLA  
 NEKAVADAMWDAFPAFTSVPSPFLGADPEAKAILRAAQDDGEGDGTASSLADSL  
 NEKAVADAMWDAFPAFTSVPSPFLGADPEAKAILRAAQDDGEGDGTASSLADSL  
 TGLRLPATLFDYPTPEALVGLRVELREADDGLGREDLREVLAAVPAEPAREKAG  
 VLDTLGLATGTGHEPTDAETTAAPAAADDAELIDLIDLSGLVQALGQIS"  
 15550..30759  
 /function="b-ketoacyl-ACP-synthase, acyltransferase,  
 dehydratase, ketoreductase, acylcarrier protein"  
 /note="ORF2"  
 PKS modules 4-6"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rifamycin polyketide synthase, type 1"  
 /protein\_id="CA11036.1"  
 /db\_xref="GI:2764762"  
 /db\_xref="GOA:O52789"  
 /db\_xref="SPTREMBL:O52789"  
 /translation="MANQSWRKMSAPNEQIVDALRSLKENVRLQENSALAAAAAB  
 PVAIVSMACRYAGIRGEPDFWRVWSEGADVYTGFPEDGMDVEGLYHPDPNPGTTY  
 VRBGAFLQDAADAGDFGFI SPREALAMPDQQLLEVSWETLREGIDPHSVRSDDI  
 GVTAVGVHDYAPDLGSGFMSLERALGTAGVAGSVAYTLLEGPAVTVDTMCS  
 SLVAHIAAQLRRGECMSMALAGSTVMTATPGFVGPARORALADFCRGSYAAADG  
 SNAEGVGVLLERLSVARERGHVLAVIRGSVNQDASNGLTAPNPSQOVRIRKA  
 LASAGLVTVDTVEGHTGTGLDPIEVQALLATYGOERDQDQPLMGLSVSVVGH  
 QASAGVAGTVLKHSHQGEIATQDPAFTQVDMVDSAGAILLEAGREWRPHRR  
 GGISFGFASGTNAHMLEEAPDEPEPTAPATVPLVVSAAATAAQLAGSAGLAEV  
 GGLVSDVAGTGLVSMALSERAVVAGSHEEATVGLRALARGESAPLISGRSGVY  
 GKVVVFPGGGTQWAGMRELLDSSEVFAARIAECETALGRWDMSTLTVLRGADIL  
 DRVTVQSPASFAVWGLAAVWASLGEPEAVVCHSGEIAAACVSGALSLEDAKVA  
 LRSAQALAAAGSAGSVALSEEDATAPLEPWAERVAGVEVAVNGTPTVITAGREALD  
 EALDLDQGVYRIRRVADYASHTHRVEAARDALAEMLGGIRAOAPVVPYTVTGM  
 VEDAGLDGCVYTRFRVFGFPAVAELIEQGHVFEVSAHPVLVQPLNELVDVDE  
 AVVTGLRRRBDGLRLLSAELVGRVTVDMNSGLPPSRVLEPVTYADHOHMLQ  
 LPTANGLVILVHAGSALRVVAPGGPDALTFOAADTGGVLTVMDSLSVRESVSAQL  
 EATAGRERDLSFOVDMTEPATTAETAHAELVLEAFGEAAPELTSRLBAVQSWLAD  
 AADEARLVVTRGAVREVTDPAGAAMVGLVRAQAENPGRIILVDTDGDVPLGLVAS  
 GEFQVLAVRGNFSPVRLARATGEVREPAPVSEFVTLTGGTGLGLVAKHLVARS  
 GQRLVLASRSGVAEEDLVTELTPEQATVSVACDVSQDQVAAALLAEHAGLQGRILA  
 GLDGDGTGIGALNRERLAGVPKVDVAQVHLDLDELTRDGLDAFVVFSSAAALMSGAGQ



[illegible]



TATGRLTDPAGCPVLTVESVRGTPFVAEGGTTDALFRVDMPEIPLPTAETADFPEYE  
 ATSAETLSALQWLADPAETELAVVTGCTEPGAAATWGLVRSQASRHPGRILVIADL  
 DDPAVLPVAVASGEQVVRNGVAVSPRLTRVTPQDARPLDPEGTVLITGCTGTGGA  
 LTARHLVTAHGVHVLVSRGEGAEELTALRGASVAIAACDVADRQAELAVLRAI  
 PAEHLPLVTAHVTAGVVDGVTTELTPDRLATVRPKVDAARLLDELREQLDAALVLF  
 SSAAGVINGQGAAYANAEALDQNSLDLPAVSIWANGVATWATVSGMTEHLGDAD  
 LRNRQIGMSGLPDAEGMALIDAAATGCTIVAARATKAGGVPPLRLGL  
 LPPRAAKTASLTARGLAEIPAALDLVRHAAVLHSGAIVSHSGRTFKD  
 AGFOSLTAVELNRLLAAATGLTSPAMLPDVPKPALADHLRAKLFGSAAANPRTGT  
 AAAEPIALVAVACFPFGCVHSPBELLRLVADGADAVTEFPADRGWDTRLIYEDPDH  
 EGTTVRHGFADNDAGDPAFFGSPNEALAMDPOORLLLETSELWELFERAIDPTTL  
 AQDITGVAGVNSHVRSMRHAAGVGFRLTGGASVLSGRVAVFVGEVPAVTVT  
 ACSSILVHMAVQALQRCESMALAGVWMTVETFEFQRGLGADPGCKAFAD  
 GADGTWGBGVGLLVERLSEARQGHVLA VRGSVNSDASGLTAPNPGSOORV  
 TKALAAAGLSVDVDAVEAHGTITLGDPIEAELATYTGONRETPWLWGSVNSNGL  
 RTAAAGVAGVITKVMAMRHGVLPRTLHVDPRESSYVDSAGAVELLTEARDVWSNHP  
 HQAGVSFGITGNHVVLEEVAAPIITTPQBPFAFLVPLVLSVARSAGLGRQAGRLA  
 AFLGRTDVRPDAAAYALATRAQLDHRVAVLASDAQCADLAAGFSGVVTGTPDG  
 KLAFLTQGSQWAGRGRELAETFPVDRDAFEAACEAVDHLRPLREVDFDGSALL  
 DOTMTQCALFAVETALPLFESWVRPGLLAGHSIGELAAHVSGVLDAAGELVA  
 AGRMLQALPAGAMVAVQATEDEVALPLDGTVCVAVNGPDSVLSGTDEAAVLAVD  
 ELAAGRKTRRLAVASHAFHSPLEPMLDPRFAVABELTYRAGSLPVVSTLTGELALD  
 SPDYVQVNRVNRSDAVTALGAGASTFLELGGGALAAALGTLGPGESCVAITL  
 RKGAEVDPVLTAELHVRGVGVTWTLDEPATAVGTVLPTFYAQHOFVWDVDET  
 KAVSVTPPBAETIPDRFQVDVLELRESAAVVLGHRDAGSFDLDRSKDHGFDLSAV  
 KLNRRLDPTGVELPSTLLFDYPNPVAVLADHLRAELLGRPAAPAPVTRDVSDEPIAI  
 VGMSTRLPAGDASPRELWKLVAERDASVGFVDRGMWLDGLYHPDPARAGSYSTRSG  
 GLVHDDYVTRLEVEDVGYTMTGASSVAGRVAYVFGFEPGPAVTPVARSQSDVGVFT  
 MHLAAQALRGCSMALAGATVMASPDFAVDFEFSQRGLSADGRCKAYAGADGTGWA  
 EGVVVVRLSVRGERHVRVLAVGSVAVNQDGSNGLTAPNPGSQRVIRGALASA  
 GLAPSDVGVHGTGALDLEVOALLATYGOEREPQPLWLSGLKSNLGHGTCAAAGV  
 VGVITKIMAMRHGVMPATLHVDETSQVDSWASAGAEVLTEAREWPTGRPRAGVVSF  
 GASGNTAHLIIEEPAAEAVDEVASVPLVVSARSAGSLAQAGLEAAVLENESLAG  
 VAGSVGRATLNRVAVTAGSRDEAQDGLQALARGENAPVVTGTPGKPVVWVFP  
 GGSGOMGRDLSSPFAARIKECAALEQWTDMSLLDVLRGDADLDVDDVVQPP  
 ASFAMVGLAAVLTGVPDVLVHSGOGEIAAACVSGALSDDAKVVALRSQAIALAG  
 ELAAGRGMSVALSDEDAVRLTPWANRVEVAANVSPSSVITAGDAQDALEALAG  
 DGVRVRAVDVAGSHRVEAETLAKTLAGIDARVPAIPVSYVLGTWISQAVDA  
 GGYNRLAAQVRFSGPSVADLAGHTVFEISAHVPLVQPLSEISDAVVTGSLRDD  
 GGRLLASAEALVKGVAVDWMTAVPAAGVNDLPTAFDRHFWLHEATEAAAGDD  
 DGEFTWIPQSDVDSIAELLEVPGEQALSTVVPVLAQRDRRERSRTEAKLYQVT  
 WOPLEAREAGVPGRWLAVPAGTITDALLKELTGQGLDVRLEIEASQAELQRLN  
 VLAHDVTLGSLIADGADAAETASTLALVQALGDTTSPALWCLTSGAVNIGI  
 QDAPVAPAAQVWGLGRAVALREDEWGLDLPALDARTAQALLGLVNGAGEDOL  
 AYRSGVRRRLVRKVPESATSRWEPGTVLVTGAEGLGRHSHASVWLQAQSERLIV  
 TGTGVDDELTAELAEFGTTVEFCADTRDAIAQLVADSEVTVAVHAADITQSSVDDT  
 GVADLDEVFAAKVTTAVWLDOLFEDTFLDAFVFSINGIWGGGGGPPAGAANVILDA  
 LVEMVRRARGIKATSIWAGALDQIGIMDEAALAAQLRRRGVTPMAPPLAVTAMVQAVAG  
 NEKVAADMDWAAFTIPTSVPSPPLFADLPEAKAILRAAODDGDGDTAGSLADSL  
 RAVPDABONRILIKLVGRHASTVLHSGAEGIPROAFOEVGFDLSAAVNLNSLHAA  
 TGLRLPATLIPDYPTPEALVGLYRLVRELLREADGLDGRDRLRVLLAVPFAFPKAG  
 VLDTLGLGADTGTGPTGDTAETTEAAPAADDAELIDLDISGLVQRLAQTS"  
 /misc\_feature  
 3883..5775  
 /gene="rifa"  
 /notes="encodes polyketide synthase loading domain"

Query Match 100.0%; Score 501; DB 1; Length 90445;

Best Local Similarity 100.0%; Pred. No. 9.1e-228;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TCGCTTCGCTGCTCGCGCTGCACATGCGCGCTGCAGCGCTGCAGCGCGCGAGTGTCTCC	60
Db	10981	TCGCTTCGCTGCTCGCGCTGCACATGCGCGCTGCAGCGCTGCAGCGCGCGAGTGTCTCC	11040
QY	61	ATGGCGCTCGCGCGCGCGCTGATGGTATGGGCACCGTTCGAGACGTTTCGTCGAGTTCCTG	120
Db	11041	ATGGCGCTCGCGCGCGCGCTGATGGTATGGGCACCGTTCGAGACGTTTCGTCGAGTTCCTG	11100
QY	121	CGCGACGCGCGGCTGGCCCCCGACGGCGCTGCAAGCGCTTCGCCGACGCGCGCGACGGC	180
Db	11101	CGCGACGCGCGGCTGGCCCCCGACGGCGCTGCAAGCGCTTCGCCGACGCGCGCGACGGC	11160

QY	181	ACCGGCTGCTCCGAGGCGCTCGGCGTCTCTCGTGGAGCGGCTGTCGAGGCTCAGCGT	240
Db	11161	ACCGGCTGCTCCGAGGCGCTCGGCGTCTCTCGTGGAGCGGCTGTCGAGGCTCAGCGT	11220
QY	241	CGCGGCGACACAGGTCTCTCCCGTGTGTCGCGGGTTCGGCGGTCACTCCGACGCGCGTGC	300
Db	11221	CGCGGCGACACAGGTCTCTCCCGTGTGTCGCGGGTTCGGCGGTCACTCCGACGCGCGTGC	11280
QY	301	AACGGCTTACAGCGCCCGCAACCGCGCTCCAGACGCGGTGATCCGCAAGGCACTGGCC	360
Db	11281	AACGGCTTACAGCGCCCGCAACCGCGCTCCAGACGCGGTGATCCGCAAGGCACTGGCC	11340
QY	361	GCGCGCGGACTGTCACATCGGACGTCGACGCGGTGGAGCGCACCGGACGAGC	420
Db	11341	GCGCGCGGACTGTCACATCGGACGTCGACGCGGTGGAGCGCACCGGACGAGC	11400
QY	421	CTGGGCGACCCGATCGAGCCCGAGCGCTGCTGGCCACCTACGCGCCAGAACCGGGAACG	480
Db	11401	CTGGGCGACCCGATCGAGCCCGAGCGCTGCTGGCCACCTACGCGCCAGAACCGGGAACG	11460
QY	481	CCGCTGCTGCTCGGCTCGGTG 501	
Db	11461	CCGCTGCTGCTCGGCTCGGTG 11481	

# RESULT 4

AF262754

LOCUS

DEFINITION

AF262754

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

1..3900

/organism="Amycolatopsis mediterranei"

/mol\_type="genomic DNA"

/strain="DSM 46095"

/db\_xref="taxon:33910"

106..>3900

/notes="similar to Amycolatopsis mediterranei S699

rifamycin biosynthetic gene cluster; contains module 2

ACP, all of module 3 including KS, AT, and ACP, as well as

part of module 4"

/codon\_start=1

/transl\_table=11

/product="rifamycin polyketide synthase"

/protein\_id="AAG16689.1"

/db\_xref="GI:10312148"

/translation="MIFDYPKPFGLADHLRGLFGENASRPBIEGTAAAVDEPIAIVA

MACRPFPGVHSPEDLRLVADGVDAVTEFPATRGMDTDLRYHEDPDHEGTYVHGAF

LDGAGDFGFIQSPNEALAMDPOORLLETSGKLFERRAIDPTTLACGDIQVFGAF

NHSDYMRTHGAAGVEGRLTGTASVLSGRVAYHFGVSPAVTVDTACSSILVALHM

AVQALQRCGAMALAGVMVGTVEFVSQRGLAPGRCKAFADGADGTGWSGV

GLILVRLSEARRRHQVLA VRGSVNSDASNGLTAPNGPQQRVIRKALAAAGLS

PSDVAEHAHGTITLGDPIEAELIATYQDREQLWIGSVKSNLGHGTQAAGAGV

IKWVAMRHGVLPTLHVDRPSHVDWSAGAVELLTEARDWTSKDHPRAGVSSFGIG

GTNAHVLEVPAPVATPRPEPAGFLVPLVLLSARSAAGLRGAARLAGFLGAEVRVPD

```
ORIGIN
Query Match      9.4%; Score 47; DB 1; Length 3900;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 TGCAGGCGTTCCGCGACGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 197
      |||||
Db 883 TGCAGGCGTTCCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 929

RESULT 5
LOCUS AB089954 58343 bp DNA linear BCT 15-FEB-2003
DEFINITION Micromonospora griseorubida gene cluster for the polyketide
           macrolide mycinamicin, partial and complete cds.
ACCESSION AB089954
VERSION AB089954.1 GI:28394148
KEYWORDS
SOURCE Micromonospora griseorubida
ORGANISM Micromonospora griseorubida
          Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
          Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
AUTHORS Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y. and Kato, F.
TITLE Organization of the biosynthetic gene cluster for the polyketide
JOURNAL macrolide mycinamicin in Micromonospora griseorubida
MEDLINE FEMS Microbiol. Lett. 218 (1), 135-141 (2003)
PUBMED 22472038
REFERENCE 12583909
AUTHORS Anzai, Y., Saito, N. and Kato, F.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Yojiro Anzai, Toho University, Faculty of
Pharmaceutical Sciences; 2-2-1 Miyama, Funabashi, Chiba 274-8510,
Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel:81-47-472-2072,
Fax:81-47-472-2086)
FEATURES
source Location/Qualifiers
    1..58343
     /organism="Micromonospora griseorubida"
     /mol_type="genomic DNA"
     /db_xref="taxon:28040"
     complement(1..910)
     /gene="mycP"
     complement(<1..910)
     /gene="mycF"
     /codon_start=1
     /transl_table=11
     /product="methyltransferase"
     /protein_id="BAC57022.1"
     /db_xref="GI:28394149"

gene
CDS
    1490..2641
     /gene="mycCI"
     /codon_start=1
     /transl_table=11
     /product="cytochrome P450"
     /protein_id="BAC57023.1"
     /db_xref="GI:28394150"

    2662..2886
     /gene="mycCII"
     /protein_id="BAC57024.1"
     /db_xref="GI:28394151"
     /translation="MRIVLDAERCVGAGQCEATAPELFTQGDGDLGLVDRPVTPELL
     GPAREAVDRCPVRAIRIESSVRTGWAGAG"

    2883..3866
     /gene="mydI"
     /protein_id="BAC57025.1"
     /db_xref="GI:28394152"
     /translation="WTRDRWANRSVVVTGALGFTGSHVEELAAAGADVLLYRSEHR
     AVRDELNDRVRLPYDLDLRLRQVFEHEASCAOTIVHCAALDGNAAVKRAHSAE
     ILDLNLRVASNLNLCVRDFGVDVTLSSAAAYCGPTASPAEDDARCTVRSNGY
     VLSKLITELAEQYRQYFGVHLVRPDRSDSFDGPARSVIPAMIRAESGGEI
     BIWGDQOTRSFVYVDLVRALALVETKFSNLNVTDTVSMLDLRAVVFVSVRT
     ARIHKPAQPGVPGADVTRMREVVDYTRTLEGTLEETVRYRHRIGR"

    3887..5086
     /gene="mycE"
     /protein_id="BAC57026.1"
     /db_xref="GI:28394153"
     /translation="MTAQTEFDEATVDQVRLIAGHDSLELRELTKYDPAMLSRLIVA
     EILSRCPSPNDTPVLVELAIVHGSERFHLRVDRDPIFVPGADEGFGVLMVYEL
     TELRELFGVTHERPAGVRGKLFPLTDDDEAVRQIGTILIAAQGTAVLAGGGR
     KPDLSLSRYFTPKFGLFWHTPHYDRHFDYRQVRLVIGYKHPGWGGSL
     RMWKSFPFRGQIYGLDINDKSHVDLRLTQGDONDAEFLLDRIARRYGFDFVDDG
     SHINARVTSFAALFFHVRPGGLYVIEDMTAYWFGGQADPQSCSTSLGLKSLI
     DAIOHQELPSPDNRSFGYVDNRNIVGLVTHNNAFVEKGRNDEGGFTWIPRDFESLVO
     ASSGGAT"

    5181..6443
     /gene="mycD"
     /protein_id="BAC57027.1"
     /db_xref="GI:28394154"
     /translation="MRIALLTIGSGDQVQPFIALGVGLRARDHVLGAPEGIKHIVE
     RAGLRYRTPDGDFFTMEPVVDSRLRSPLYRLDLLKALPAVEEDQVVDIEAAG
     QGADAVVYAPLITIAAYGRSDVFAAANWNTWRFPFVSGQRRLGSLPLXHLF
     THARAREWARRPVDYRRQGLFPGRSPFLRLGRPPYLYPFSPVLPKPRD
     WPARAHVYGVFWDDQVPTRELVDLADGAPAPVADYRQHLAVEREOLVYATAAA
     RAAGRRIVVDGDPDVPDSDVLRLDHVDYRWLFPTAAVHHGGGTAAERVVGRPH
     VYVPVFAHPHFWARLTETGVAAPAVDFTLTREALVASTVKAVEREDSLGARAARLGE
```



```

/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translation="MLGHRPSGILLIVRYVAQALRKKLAEKSTADRHNPSSGAPM
ARLTPELPGASTREVDIVQVNLNAGYAGRSQDDVAHIAELAEIGVPTSPYALYPVA
PYLAQIDRVAVQHRRTSGAEALVAGGEGELLTACDHTDRDLDEVHGVAMSKNAG
PDVLAARWLADVEPLDILLTARWTRDTGETEIOHGTIAELLTPAYVWDVLSRD
ALTPGTVLISGTTIPMTPEGVQDFADTWRELVGDEATGDTIRLAYDVHMPPEPIG"
3117..3818
/gene="nanT3"
CDS
3117..3818
/gene="nanT3"
/translation="MSTRCTGSLTNGWWSVELGIPLSVLVDHPVVRAGISAWCAA
DAPISVVAEGANVSALYHGPGRGADVVMDLILQNGRPAYDELQELVAQERKVVYTM
RDSQDAALTCMDLGSATYITKABGQRHLVKAIIRAAEDIPYTPSLAGAFGSDTROSR
PVLVSREVEVLVWFQSESKAVVAQSLGISERTVNTVLDVRVRIKYANAGREATTAKL
VARAVQDLTALDEL"
3117..4923
/gene="nanT4"
CDS
3117..4923
/gene="nanT4"
/translation="MGEDPQLITRHOIRAGGLVILVTLHGIALPGLGLGLEY
RHWPIAAVGLTLILGGSVILGLDGRPMWPCALAGTVFASVATTSQVPADHY
FVSLSSAGSLAGFWVLLGHGRLVSGAFGLAHVATYALLTLVDVPSRSIGASMA
LSALSACFQNTIAGAKLLDSALQALRAQERVTRIAVARQIQADQRRRIYEL
NATVPLTLGLAGSLDPEDEVRHACALEARLRLFAESDCTFPLVHEMRACIDV
AENGTQVLAVRGDPLEPLRLRALLDPVIAALAAAGTARTVTVRGGDQVRGVV
VDALRDLPEPKANGIRVTRVHAEDRLVAAACRIARTARPELSDRPEPPWPPWSP
AGRRSRT"
3117..6106
/gene="nanT5"
CDS
3117..6106
/gene="nanT5"
/translation="MSISLSLSPGIDPPASARLPSSAGPWAREVYVFNRLALYGRAGV
MSACVLGVAMPMEGAPAAVAAVLAWSLTHRLATTSAMRQRLPLDALDVTMTA
LCLSQRETIPEAGTTHAGTWILVSVFTAVAYGMQPLLGLILLTLWCDDVVGTL
HPNLESGVRSVWVNTALAKARVRLVSESRADEADAARARQQGEAEARC
AAERHLAHLDTACLTLLTASAPWASMAETLRQAARDLLRLRAEQPVVGLAA
ELUDETAHPLRVINRVEBLGTTWRPIAALRGGLGEALRNVARVAGVDVAVTAE
VDHVIIVTADSGVGDPDEPHIPGHIGIKRSIHGRMWNVGGRAVESRPGHGTVRL
WPRG"
3117..7493
/gene="nanT6"
CDS
3117..7493
/gene="nanT6"
/translation="MGPLRVVDEKGSPLRAQKLAALVAILLRSQVYMAEOLTREI
WGLPRPRDPASIVCISRLRKLRSRQGNPLITPQGMFLRRASCECDPFRVQQLV
ADGRKVCYRGEPAQALCFQALALMRGPAFGVGTGPIVAGLKLRLDEIRCAEMS
AEAHLEIGHREHLVGRLSLVAFPPURFETFYRQIMLALYRSQKSEALGIYESARVL

```

```

DRELGLPGRKLRRLRHTVLA"
complement(7573..8334)
/gene="nanR2"
CDS
7573..8334
/gene="nanR2"
/notes="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR2"
/protein_id="AAP42854.1"
/db_xref="GI:31044142"
/translation="MRYEILGPLRVVDRSSVTSISKIEVLAAALVRADEVSSD
LVKEIWGTAPRAQVAAALHYVSQLRFLKRPNRTSPVITARSQVYRLGCDDELDFH
VFOGLVQGRRAQAGRAGTHACAEYEGLEFWRGPVLDVAGVINGFCVWCAQARL
ECVESFEAGLEIGHREHREFVSLYARIEELPLNEAFYRQLMLALYRSGCRADALHYVQ
SAPAVLSEELGLEPSAALKRLHVALVDAPELSAA"
8919..17627
/gene="nanR3"
CDS
8919..17627
/gene="nanR3"
/notes="PKS"
/codon_start=1
/transl_table=11
/product="NanR3"
/protein_id="AAP42855.1"
/db_xref="GI:31044143"
/translation="MRFRYGVNAGSSPSHAQCATSPVAIVGLACRLPGADPEAFWR
LIRAGENAVVPDSDRLPTPEGSPYFAGLLEHVDVDFDAGFGISPREATAMPQORL
MLELAWEMEDAGLPKNLAERTVFTGAIWDDIATLLHRRRPNDIAITRHTMAGLH
RGLIANRVSHLLGLRPSLTVDAASGLVAVHLCESLRRGEADLALAGCVNLILAE
ESNRMAEQFGLSPDCRCYTFDARANGFVREGGWLKLLAAAVADGDPVYCVIE
GSANNMGATDGLTRPSADAQTDVROAWORAASPAELOVELHGTGTVPVGPPIEA
ALGDALGDAGDCHRDRAVDAPLRVGSTMVNGHLEAAAGIAGLKTALSIHRRLLP
SLNFTNPICPLAELGLAVQAFAGPWPERRRLTAGVSFGMGTCNVHVLAEPPAP
AVLPDPTAPGVSAITAPFPWPVWSAASPKALTAQAALYEHKRAHGPGLHVPDIGH
ALATITPRAVAVILGRDEDELVSRDLALASSETQTSQVIRGAGGRVAFPLFGQS
QREGMRELYAAYPVFADALREVCALHDMPLDITDPLDLMAFAEPDGPFPPLNRTAY
TOPALPIVALYRLVTSWGTDPHLMHSHVGEITAAHVAGVLSLPDACLVAARGEL
MOSITAPGMAAQAATAEAGQALAEVGGVGLAAVNAPAVVISGDREAVATAEAW
RAGRKATVLYKSHAFSLHDLGLDLRTVAAGLTPAALIPVNSLTGGAATEAQL
RSPDWDHARQAVRFDAGVRHLCDAAGVDTLELGDPSALUTGWARSAAGADAPRL
VAVVQRGRPEAQSFVSAQAQVHGVGVDAWAAAFAGHETRAPLFTYAFQDRHWP
DGLDERGARRPSTPVVSTDRFPVVDASPADRAASPGELLALVTHAALVLGHNSP
DGIDPALTFKGLGFDLSLAELSERLSAATDTDELPAITLPHDPTNAAVLAARHEG
OPTAAATAATGSMEDPVAVVAVSCRYPGVESGEALNELVDEGVDAVECEPGRGW
DLAELGRAPDCCSGSATCGGFLYCGAPDAEFPGISPREALAMPQQRILLELWE
LIERAGIPASLAGSATGVTVGATVYDGRHEATAEELDHLLTGSTPSVSGRVAY
ALGEGPALTVDITACSSSLVAMHAAQALRQGECDLALAGGVTMATPGMFTSFRQY
GLAPDGRCKPFAAADGTGNSGAGLVLRLSDARRNGHQLAVIRGSANQDGASN
GLSAPNGSQQRVITROALANARLEPADVDAEAHGTGTLTGDPIEAQALITATYGGORT
DRLPLGSIKSNIGHTQAAAGVAVIKMVALRHGRLPASLHIDAPSPHIDWSGTV
RLLSEVDPGTDTPGSDRPRRAVSVSGISGCTNAHLILLEQADPHPEPTTSGVVP
WLSARTALRAQAQRLAEVWTAGAPSPASPSPASPASPASPASPASPASPASPASP
VSGTDRELLSGLRAVDGLAPAAAGAPGPMVFPQCGQMRGMGVELLDSSPVF
AARMACEALGELFVMSLTVLARGAPAPESRVDVLPCLMAVMSVLAASVWESYGV
TPTVAVHSGEITAAACVAGLSLRDQARVVALRQALRALAGLMSIALSGAE
RFLADIAGAAARVAVAVENGVPSTVSGTDDVAAVVAACEAGHARTIDVYASR
POVDRLADITRIDLADLSFGASDAVYSVARGQTEHLDADYDFTNLRQVRFASA
IDLAAAGTRVFEVSPHPLPALRECEEEAAVAAATVTLRRDDGGDPQVVALGD
GFVALGVADWSRMFVGDEGREGDEHRTVELPTYPQRRYRLWADGGPDRREGRTAG
VGTTPAGHALLSASVELADGGLVSLGRPLGDAWVGAHTVAGVQLVPGALVDAWLLIA
ADBAGGASLELLRAPLBSGLSPSEPSAGVAVOAVAGPDESRELRITSSRAD
AGAGCWTCHVAGSLAPGPPADPTGTATVPWPPAGAEALDPAGIYERAEPRGYVG
PALRGVWLRDGLADVADVLPPEERAGGEGGADGTAGLHVLVDLDAALQPLL
APEDGTGGGAGGAEKRLWLPFAWGSVRLWATGARAARVRLPLDGGGGVDAARELRI
EVSDDTGPVLSVAVSLRPTVVRQREASGAAGLFDALDTPWPAQEPGSGEDDAG
CVAVLGEAPTEPGDRCRTYDTPALLAALDAGAPLPSVWMPRPAQEPGAAPDEAA
LSARGVGAALRAWVAERPLTVSRILAVTRGVAAGGEPVDAALAAAGCARGVQ
ABHPDRILVLDVDDVDMCADTDTDIGAANGAALGEPQVVALRGDTLLAPRLASAA
TPGGVAFDNGTVLIVTDSGSLAGSVAEHLVRAEGVRHLLVRFEGADGADYDIDRQD
AQVHWVTPDRTOALERVAQVDPFAHPLTVGVHVAAGLSADTETSAGRGVAVAGV
RALHQATPALPSVRFVTLSDATANDGPAAPEAAAGAFCAAVTVDRRAGLGLDVA
FGWAAADDDGGADSGRWNTGVLGADRGLALLRAACRADRPRLVAADIRTRALTAHPA

```



APDAAEDEAAGSGTVPALLSGRSASALRAQAARLLSRLLDGDGPGRITDVAISLATOR  
 SAFPHRAVILANRADLHLSALAEAGTEAPVAVOARASGKLAFSFCQSORLG  
 MGRELXGYPFAFALDAVCAALDAHLDRDVLWDIWDGDEALLNRTGVAOTGLFAIEV  
 ALFRLLSNGVVRPHLLGHSIGETAARAHVAGVLSIPDCAALVAARGRLMQOLPSGGM  
 MAIRATEDVPLHAEGLSAAVNGPSSVVGSAEDEVIALAALHPEERGRKTRIRVRS  
 HAFESPLMEPLMADFRFRAVADGMYTAAPRIPVSVNVTGRPATAEELCCRAYWVGHREA  
 VRFADGVALREQVTTLELPGDGSLSALAAESAADDVLAFLVRKNRPAPALLTA  
 LARLHAQGTVDWGAARLTGARWDLPTVAFQHERFWPSCGAARAGDVRGAGLSGAG  
 HELLGAVELAGSGRLTGRLSLSSHPLADHVVLGSLVPGTALMELVLRADAEVD  
 CAANDELTLAFLVLPASGAALQVQVWGEDEAGRRVSVHAREGEGPETHADGAL  
 APAATVPFDTAIWPFGAEDHDAAGCYERFADAGFAYPVFGQLRAAWKLGEDIYAE  
 VALPEGTGNGVHPALFDAAALHAALGSGGTDEAAVFPFQNGVTLTHAGSRVVR  
 IRPTEGTSIALVDTAGPVAVSRLTARPITAGLOLQDGRDLSLFQVDMTTLHTDER  
 ANSLALCKGTEDLDTLSLOPHADLDLAATGVHDTVLALPLTRTACTGVBESVRAAT  
 GALALIRSLADDRPAAARIYVTRGAVSGTDLGASVWGSLVRSALLHHPGRGLVDV  
 DVQDAEVLVPRALASDEPQVLRGGEVLAARVRAQSSVTWDPGTVLITGCTG  
 GLGSRVHLSEHVRSLILVRSRGPAAEGVDALVAELAECAQVTEACDVTDAVA  
 VADLVARRIRISAVVHTAGVLDGCVVESITPERLSAVLRPKVDAAWNLHETGGLDLD  
 FVFSVAGTFSAGQANYAAGNFDLALYHRAVGLPAVSLAMGPSQDGMGTGL  
 SDADVORTARQMPPLTVEGLALFDAALGSAEPMALPVRLDLAALRAQGPOLLRG  
 LIRTPGRTAAAETGDTAAAFAGRLTGLSAGREGLVLAARSQIAGVLGHABATEI  
 QDRAFLDLGDSLTAVELNRLGAVTGIRLPAITLFDYPTPAELVAHLHARLAPPT  
 VGPALLGELERMEKSGGLDTEENHEHQIAGRLVLRKAKMDALRDTAAAGHDGSPS  
 DEDEFESADDEVEDLIDNEILGUS"

## ORIGIN

Query Match 7.2%; Score 36; DB 1; Length 20394;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 CGCGACGCGGCTGCCCCGACGCCGCTGCAAG 156  
 |||||  
 Db 5320 CGCGACGCGGCTGCCCCGACGCCGCTGCAAG 5355  
 |||||

RESULT 8  
 LOCUS AX067996 20394 bp DNA linear PAT 19-JAN-2001  
 DEFINITION Sequence 3 from Patent WO0077222.  
 ACCESSION AX067996  
 VERSION AX067996.1 GI:12329806  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Streptomyces natalensis  
 Bacteria; Actinobacteria  
 Streptomycetaceae; Streptomyces.  
 1  
 Martin, J.F., Aparicio, J.F. and Colina, A.J.  
 Genes encoding enzymes in the biosynthesis of pimarinin and the  
 application thereof  
 Patent: WO 0077222-A 3 21-DEC-2000;  
 DSM N.V. (NL)

FEATURES Location/Qualifiers  
 1..20394  
 /organism="Streptomyces natalensis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:68242"  
 1..20394  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC24642.1"  
 /db\_xref="GI:12329807"  
 /db\_xref="RENTREMBL:CAC24642"  
 /translation="MNEKRLRYLKRATADLHETRQQLDETEAKOREPLAIVSMACR  
 FPGVRSPELWELLRDGVDVSSFFPNRNGWDLALYSDPDPAHQGTSYAREGFLHDA  
 GFDPFGFISPREALAMPQORLLETAEWAVERAGIDIPESLAGQSTGVYGTGHHG  
 YDAEGRRADEVGCHLLTNHISIASGRISYVLGLEGPALTPAAADGSGVLSLHAMH  
 ALRRDCRAALVCGATVMSTPQVFRQGLAPDRCKPFAAAGDGTGWSRGVLL  
 LVERLSDAVRNGYPVLALVKGSAVNODGSNGLTANPGSQQRVIRQALTGLAASD  
 IDAVEAGHTGLIGDVPVHAALLATYQQQRAADRCPGLGSMKSNIGHTQAAAGIAGM  
 KMVLARHGHGLPTLHLDFTGVDWSEGNRLLAEPWPFPSPRRRAAVSSFGISG  
 TNHVLLEQAPEAEAPAPARPALPILSARTAEAGLRAQDLRGLRDRADLE

## CDS

PAVAHALADTRTILMEHRAVVVAGDREELRGLDALLAAGRANGLVSGVAKAAAFSL  
 FAGGSRPGMGRHELHAAHPVFAFADVCAEDELPHLDRLDRDVFVFAEESGAALDL  
 QTYATQALPALETALFRLVESWGVAFRFVAGHSIGELTAHVSGLVTHDAAARLVAA  
 RGTLMQALPAGAGVAVQATDETRERLAGHEDHVALAAAGPDSVTLSGDEQAVTEI  
 AAHWEAQGRTRKLRVSHAEHSFHMDDMLEDFRVARGLTFHAPRIIPVSTVTGALAT  
 EDELRSDPYVVRVRETRVFCAAVTRTILAEAGVTTVEIGTGGVLTIPMVGDCLTLEEP  
 VLPLLTGRPTVALTEGVAATVAVHGVVDSPFAPPGTSGRDLPTIYAFQRMVWL  
 DPADHDEGAARAAEAGEAGFWAVERDQLSALVALDGEASLSGLFTLSWRR  
 QRTTAAADRSFYRTHWAPRTAGPTATGHLVVLPEGGTDDPTARLLDALNDQGL  
 HTDRELPADEHPDAGRHVPDGVLCILALDERTRSCPYPVRRGLAATNAARPEGGA  
 GIQAPLWCTVGAVAVDRHEALPQAOQWGLGRVALESQSGWGLDILDPNLDGR  
 AVSALLTAGEEDQVAVRPAVGFARLERITPGDGTGDRMSTHGTVLVTGGTGAAGA  
 HLHAILDAGAEHLVLTGRRGPQAPAPELAAALTDRGVKVAALCAADRDLALHVL  
 ADIIPHLPLTGVHAAGYLDGVDLALTPEFETVLPKRAQAQNHLELTQDLDDHF  
 VFGSIVGLGNAGQANYAANAALADLAHLAQGLPATSPVSGWGGQAAAHWSDAA  
 DMRSGLGILPMAAARRRPPAALAQMGTVTVADIDWSAYAPALTAVPSPILGDLPE  
 ARRALPAGRPRERSPLDRIGALPPAEQKAFITWVREBAARVLGHPSPTVDQAR  
 AFRQGFDSLMADLRLNSAATGIRLPLTLFPHPTLALAAACRLSEVLGAGATPV  
 VOASTAALDEPVAIIGKARFPFGVHSPALWRLIAEGGDALTPMADGMDLRLYH  
 PDPHQGTSYARGGGLDGAADFADDFGIISPREALMDPOORLLLETWELWEGAGID  
 PESLRGSGTGFAGTNTQDYGTLDAADQDEAGHRLTGNAMSVSGRSTYTFEGEPGA  
 LTVDTACSSILVALHMAAQLRQEGCSLAVAGGVTMATPGSFVFAQRGLAPGRC  
 KPEAAADGTGWSGVLVVERLSDARENGHOVLAVRGSVNODGASNGLSASGP  
 SQQRVIRQALANARVAASEVDAVENHGTGTLGDLPIEAQALLATYQORFLLGAVKS  
 NCHTQAAAGVAGVMKWLARHGMPLRTHVDEPTGHVDVTAGVATEHTEDMPTET  
 GHPRRAVSAFGISGTNAHVLELPAEQPLVEQPSAAEPDAPATPRTPTASDGT  
 PLLLSAKSESALRAQARALHSHLERDPAALRTDAAATLMTHTRTAFHRAAVRAAHEA  
 ALRALTALAAGEADPAVTGTHTGRDAVLFSGQSGQRITGMRELSGRYPVFAEAFDT  
 VCAALDEHLDRPDRVGRDEELANRTVYAAQGLFAIEVALFRVLESWGVPHVYVAG  
 HSGVGLAAHVAGVFSLADACALVAARGLMQALPAGGMARITAEDEVLPHLADS  
 STIAVGPSSVVVSGAEHAVLSIAAHFSGAGKRTLRVSHAFHFLMDPLMADFRAY  
 ABLGTYGEPELAVSNVTGQLATDQLRTPPEYVTHVRAAFAFGIRALGAEGVTRFL  
 LELGDPGLSALAREAPDADACTPVLNRDRSEAAATLAAHLTHVHGTDEMTDPTFA  
 GRDAHDLPTVAFQHFPTWPTDTRTGDICAVGLEATGHPLLSAAVLEPDEGELLF  
 TTRLSLOTHPLAGHVMSVLLPCTAPAEALAVLEADGVCDRDELTAAPLVLEH  
 GGVLQILVGPADASGRITLARSAGGDGDPWQHATGVLAEGESTPEGCVGTHTE  
 SNPPADAADSLGLYPDFAHGFDPGFHFGQLRTAWRGDDEVFAEVALPAAEGEAS  
 AYGLPALLDLALHVVAVGDRVGVFPFSSWESVALHATGASAVRIRVHSGDTSVD  
 VADTTCEPVASITGLVLRASADQLAGGADPAVDALFRVONVPLPAGAAVTVAT  
 LGLSAGPEFDPYDLASLRGAVAGVLPVEAGAGEVADVVDVVGATHAARALDL  
 ARSLWADDRFAASRLVFTRGAVSGADLAGAVGLVRSALSEHPGRLGVLDDDDAE  
 LALVPRVLDEPQLLVRGGEVLAARLARAQSSHAVTWDPSCVTAVTGGTGGRMA  
 RLHVHGVNRLLVSRGPAAGAEELVTELRHSGAEVAVACDVTDAAAVADILVAR  
 HRISAVHTAGVLDGCVVESLTPERLSAVLRPKVDAAWNLHETDLDLDAFPVSSV  
 AGTIGSPQANYAAGNFDLALHRAAGLPAASLAMPSPDCMGTTGLTDVSSA  
 SPGRHARTHPTRGRLFDALAAAGDAHLIPVFWASRAQGEVPLLRLLRTRARR  
 SAVGSAAGAVGLRSGRTVERREVLIVRAQILAVVLGHANPETTISTRVFDOLG  
 FLSLTAVELNRNLNATGLRLSATAVFYPTADALVDLDELFGAQEAEALPAPVPS  
 PAGAADDPVVIVGMSCRYPGGVSPEDLNRVLSGVDVADVSDFTRGMDVESLISDPP  
 SILRGSTGVFAGVMTYDGLLVGQDFEGYRSGSAASIASGRVSYTFEGEPATVVD  
 EALGTSYTRSGGFLHEAAFDPDFFGMSPREALATDAQORLLLETWELAEIKTIDPA  
 TACSSILVALHMAAQLRQEGCSLAVAGGVTMSTPTTFVFSRQRLSAGDRCKAF  
 DAADGVWGEGVGLVLELSDARENGHVLAVRGSVNODGASNGLTAPNPAQQR  
 NTHQAASAGSVAGVAKMVLAMRGVLPRTLHVDEPSSHVDMSAGAVELLSAEWPQG  
 EGPFRAGVSSFGVSGTNAHILEPQPDADAAPATVTDPGALAVLSARNEAALURC  
 QAARLLSLVAGSDALCARDIGHSLVTRSGSFARVAVWQDRDALVRALSALAEAD  
 AGLAEGSAGARTAFLESGQSORGLMGWELARYVPFADAFDVAALDEHLERPLR  
 DYVWGEDAELLNQTAYAAQGLFAIEVALVLAESWGMPPDFVAGHSIGVAAHVSGV  
 FSLPACALVAARGRLMQOLPSGGMARITAEDEVLPHLAEGVSLAVNGPSSVVIS  
 GAEVDAIATAAHFAGEGRKTRLRVSHAFSPMPLEPFAVTRLSFGTPTIPV  
 SNLTGLAPEQLAHADYVWRHREAVRFPADGIQALRAEGVTRFLELPGDGVLSANAR  
 ESADDDAVLAPVLRDRPEETALLGALQLYRGHVDWTVFPFAGSGARWADLPTVAF  
 OHERFWPSPGVARPGDVRGAGSLGAGHPLLGAALVLAGSGGLFTFGRVSSHPWLAD  
 HVLGSLVPGTALVELVLRADAEGCDLLELTAAAPLVLPASGAAVQVQVAVGPRD  
 EAGRFPVSHAREGEGPWLTHASGAVTSCEVPPDPAATVMPKGAEPVDVQCYDVL  
 DAGLITGFAGHGLQAARKLGGDVYAEAKLPSTGDGAYGLHLPALDPAALHASALGAE  
 AGCVFPFSWAGVLSHATGSHLRVIRREAGGALSVIADTSGAPVAVSVESVIRPLSAG  
 QVQADRDALFKADVPVPLTDERVEPGTGPGEPLRTYDADLDSLEGAAVPGTVLAP  
 PSAGAGTVESHAATVWALEWQVAMDRTFATSLRVFVTRGAAFGADLAAAVRGLV  
 RQAQSNPFERFGLVMDGDADTVPQAALATPEPELLVRGGEVLAARVRAQSSHTVT  
 WDPGSGVLTITGGTGLGRSVARHLSVHGVRSLLVSRGPAAGGELVLELRGSGA  
 EYVIEACDVTDAVADILVARHRIISAVHTAGVLDGVDGVVESLTPERLAVLRKVDAA



WNLEATRGDLDDAFVPSVAGTSGAGQANYAAGNAFLDALAYHRRRAVGLPAVSLA  
WGPWQDGMGTSLSDAIVARIQGMPLTVEEGLALFDDAALGSAEPMALPVRDLA  
ALRAQGEQPIRLGLRLTRRSAGAAAGSLAQLAGLSTAEERREALLDVVRAQIATV  
LGHAGPETIADPRAQDGLDGLSLTAIRLNLNKGATGRLATVTFDPTVDALAAHL  
LDELFGAETGATETPTFVPLGSLADDFSVIWMSCRFPGGVSFDPDLRLWLDGVD  
AVSAPFTDWEIIDDYDPEREGATATSGGLFDDADEFDFGMSPREALTITDAQQ  
RLLLETTWEALREAGMDPATLRSRTGVFAGVMDHYDSTLLSGFEFGYQSGSAGSV  
ASGRVSTFGPEGPAVDTACSSSLVALHAAQSRGCSLALAGGVVMTSTPLTF  
VEFSQGLSADGCKCAADADGVMAEGAGILVLERLSDAARNHRLILATVRCGSA  
NODGASNGITAPNGPAQRVIRIQALASAGLSADVDVAEAGTGTGLDPIEAQALLA  
TYGQRPEDRPLLLGSVNSNIGHAQAAGVGIKMWLARRHGVLPRTIHDVDEPSSHV  
DMSGAVELLTSEAPGEGERRAGVSFGISGTNAHVILEPQEPVAAETESITPTD  
APDAEADAAAGTVPALLSGRSALRAQAALLSRIDGDPGRITDVAIYSLATRG  
SAFPHRAVTLAANRDLTSLALAEHTTEAPVAVQADKARSGKLAFIPSGOGSRLG  
MGLRLXGRYPAEALDACAALDORPLRDVWGEDAELLNRTGTAQTCGLPAIEV  
ALFRLESGVPRDHLHLSIGIEIAAAHVAGVLSLPDACAALVAAGRLMQOLPSGGM  
MALRATEDEVLPHLAEGVSLAAVNGSSVVVSGADEVLAALAAHFEERGRKTRKVS  
MAIRATEDEVPLHAEAGVSLAAVNGSSVVVSGADEVLAALAAHFEERGRKTRKVS  
HAFSPGLMEPLADFRVADGMVTAAPRIPVSVNTGRTAPABEILCCABYVWGHREA  
VRFADGVGALREQGVTTIFELGPDGSLAALAESADDSVLAFLVRKNRPEAPALLTA  
LARLHAQGTTPVDMSSAFACTGARWDLPVAFQHERFVPSGGAARAGDVRSGAGSAG  
HPLLAGAVELAGSGRLITGRSLSSHPWLAHDVVLGVLPVGTALMEILVIRAADBD  
CAAVDELILALPLVPASGAALQOVWNGEPEAGRPVSVHARSGEGPMTLHADGAL  
APAAETPDDTAIWPPQGAHLDACYERFADAGFAFGVFPQGLRAAKWLGEDIYAE  
VALPCTGDNAYGLHPALFADALHAALIGEGEDDEAAVFFSWNGVTLHATGASRVVR  
IRPTGEGTIALVDTAGAPVASVRLTARPTAGQITGDRDLSLFTQVDMTTLHLTDER  
ANSLALGKDTGEGILDTLSIOPHADLDLAAATGVHDVTLAPLTATLHAGTVEGSHAAIT  
GALALRWLADDRFAASRLVFTVRGASGTDLAGASVGLVRSALHHPGRFGLVDV  
DVDQAEVPLVRALASRLVFTVRGASGTDLAGASVGLVRSALHHPGRFGLVDV  
GLGRSVPLVSEHVSGLLVRSGPAAAGVDALVAELAECAQVTVREACDVTDAVA  
VADLVARRHSISAVTAGVLDGVSLETPERLSAVLRPKVDAANLHETRLGLDLDA  
FVFSVSGTGFSGAGQANYAAGNAFLDALAYHRRRAVGLPAVSLANGPNSQDGMGTGL  
SDADVIRIAQAMPPTVEEGLALFDDAALGSAEPMALPVRDLAALRAQGEQPIRLG  
LIRTPGRRAAATGDTAAAFAGRLTGLSAAEGREVLGAVRSQIAGVYLGHAEATEI  
DQDRAFLDGLFSLFRAVELNRILGATVGRILPATLLFDIPTPAELVAHLHARLEPPT  
VDEDFPESASDEVEFDLLDNLGLS"

ORIGIN

Query Match 7.2%; Score 36; DB 6; Length 20394;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 CGCAGCGCGGCTGCGCCCGACGCGCGCTGCAAG 156  
|||||  
DB 5320 CGCAGCGCGGCTGCGCCCGACGCGCGCTGCAAG 5355  
|||||

RESULT 9  
SFU78289 43280 bp DNA linear BCT 13-AUG-1997  
LOCUS Streptomyces fradiae tyactone synthase, starter module and modules  
DEFINITION 1-7, (tylg) gene, complete cds.  
ACCESSION U78289  
VERSION U78289.1 GI:2317859  
KEYWORDS  
SOURCE Streptomyces fradiae  
ORGANISM Streptomyces fradiae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 43280)  
DeHoff,B.S., Sutton,K.I. and Rostek,P.R. Jr.  
TITLE Sequence of Streptomyces fradiae tyactone synthase gene tylg  
JOURNAL Unpublished (1996)  
REFERENCE 2 (bases 1 to 43280)  
DeHoff,B.S., Sutton,K.I. and Rostek,P.R. Jr.  
TITLE Direct Submission  
JOURNAL Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate  
Center, Indianapolis, IN 46285, USA  
tylg.  
COMMENT Location/Qualifiers  
FEATURES  
source  
1..43280  
/organism="Streptomyces fradiae"  
/mol\_type="genomic DNA"

/db\_xref="Taxon:1906"  
1..43280  
/gene="tylg"  
816..14234  
/gene="tylg"  
/note="ORF1; putative"  
/codon\_start=1  
/product="tylactone synthase starter module and modules 1  
& 2"  
/protein\_id="AAB66504.1"  
/db\_xref="GI:2317860"  
/translation="MSSALLRAVQNCVGDIMTNTAAQNTQDQEDVDVDPSTHGGE  
TAVGMSCLFGAAGVEFWELLRSRGMPTQDDGTWRAALDEHAGFDAGFPMNAR  
QAAATDPOHRLMELGWEALDAGTVPGLTGTCTGFAGVADSDYAVLTRSASVAG  
GYTATGLHRLAANRLSHFLGRLPSLVDSQASLAVQLACSRLRGETSLVAG  
GVNLLITESTTWMERMGALSPDGRCHTFDANGVVRGEGGAVLKPDLAALADGD  
RYVICKGAVNNDDGGASLTTPDREAQVLRQYRRAAGVSTGAVRVVLEHGTGTRA  
GDPVEAALGAVLGAAGSGRSTPLAVGSKVTNKGHLEGAAGTGLIKATLCVRKGE  
VPSLNFSTPNPDIPLDLRLRYTEROEENEDDRPVAGVSFGMGSLTNTVHLVIAEA  
PAAAGSGAGSGAGSAGISAVGVVPPVVSRSRVRVREAGRLAEVREAGVGLA  
DVATMADRSGFRYAVLARGEAEALGRALAGDGDAGVTVGAVLDGCVVVGAAAP  
CGAGAGGAGAGAGAGGGVVLVFPQGTQWVMGAGLGSSEVFAASMRCEARLSV  
HYWDLLEVVSGAGLERVDVQVPTWAVMSIARYWQMGVDVAVVCHSOGIEIAA  
TVAGALSDEAAVVALRAGLIRGLAGGAAVPLPAGVEVREAGLAKWPGVEVAVN  
GPASTVSGDRRAVAGYVAVCAQEGVQARLIIPDYASHRSHVEDLKEGLERFISGIRP  
VLRVAIEOTAEAADESGEPFDAGYFNRRLNRVFSVAVGGLLEHRRFTFVSAPH  
TTLTYPFNNHHWLDITDGGGDDATOEKESPLTRELRLGSPSSOKGLFLDLVCR  
HTAVLHETLDAABDFDLSFKQKQISMTGKVBELRNLLTETGLAETIYVDPPTPRA  
LAQVILTELLDGGSPSVLAQAQKFAEQPIAVVGMGRFPFGVSPSALMRLVVEG  
VDVAPFPDGRGWDEGLYDPEPVAGKSYVREGGFLHDAEFDFAFFGIGSPREAVM  
DPQOKLTWEATIERAGIDPHSLHSGRTGVAGVQPEYGFRLAEGDSGSLYT  
GTFSGVSRVATLGLGPAVTVDTACSSILVALHLAVALRGCDMALAGGVTVM  
AGPGEVFSRORGLAACRCYAFADGADGTAAEGAGVILVERLSDRRLGHPVLAV  
VCGSVNQASNGLTAPSQSPQVIRIAGLNARLTIVADVVEAHGTGTRLDPIE  
AQAALLTGDRDRDGGPWWLGSNIGHAQAAGVGIKMWLARRHGVLPRTIHDVDEPSSHV  
EPSSVNDVSGAGVWLLTEARFPGVDRPRAAVSAFVSGTNAHLILEAPDTAEASA  
TTPVRSVSESAVLDARSGVVPVVSRSRVRVREAGRLAEVREAGVGLADVAVT  
MAGRSFGRVAVVLARGEAEALGRALAGDGDAGVTVGAVLDGCVVVGAAAPVLF  
PQGTQWVMGAGLGSSEVFAASMRCEARLSVHVWDLLEVVSGAGLERVDVQV  
TVWAVMSIARYWQMGVDVAVVCHSOGIEIAAATVAGALSDEAAVVALRAGLIR  
YLAGGAAVPLPAGVEVREAGLAKWPGVEVREAGVGLADVAVT  
GVQARLIIPDYASHRSHVEDLKEGLERFISGIRPVPVCSVAGEQSPGFVQAG  
YFNRRLNRVFSVAVGGLLEHRRFTFVSAPHLVHAIETAEADRSVHATGTR  
RODPSRLHLLTSTAEAAHAGTLDWDPALPGHLLTTPYFNNHHWLDITTPPTAT  
TTQSPDANRYRVTKALTESPVRPHSICRLVAPPTTDDGLLGLTSLSERGAS  
VARLEPITGARRAEVABLKPSSMESAGEENTVSLILGLVPSDVAITSALLQAVSD  
IGVPAARVVALTRRAVAVPGETPDQAGAQWGRVVALELPDIPDLWGLLIDLPETABL  
TRTPETSQPPQTPERLPQTPNRRALAAVLAGRDGQVAVRAGIYGRVSRRAA  
AGAAQWQPGTGLITGGMGAIGRRLARLAAEGARLVLTSRRGPAPGAELAEELR  
GHGCEVHAACDVAREDALAAVTPYPAVNAFHTAGLDDADITLSPESFEVVRGAK  
VCCAEILLHQLTADIKGLDAPVLFSSVTGTWGNAGQAGVAAANALDALAEERRAAGLP  
ATSVAMGLWGGGGAAGAGEESLRRGLRAMDPDAVDALLGAMGRDVCVTVVDVDM  
REFAPATNARIPGRLEDTVPEAREALTAAGTTSTPDGAPELLARLSMLNETBLRKL  
VELVTEAARVLRHNTDAIGARFPKSGFDSLTSLERLNLNAGTGLKLPATVIFD  
HPSPTALARILLRLTGAGAPADAPPLPVAADDDPVIVGMACRFPFGAGTPEA  
LWKLVTREOVIGAAPTRDGMGLDSDVYDPEPVAGKTYVREGGFLHDAEFDFAFFGIG  
SPREAVAMPQORLLLETSHWATERAGIDPHSLHSGRTGVVVGITHQEYASRLHEAPE  
EYEGYLLTGKASVSVSYTIGLEGPUSLIDTACSSILVALHNAQAALRGCECDMA  
LAGCVTVMAAPGLFVFSRQGRKADGCKAFADGADGTAAEGAGVILVERLSDRRL  
LGHPVLAVCGSVNQASNGLTAPSQSPQVIRIAGLNARLTIVADVVEAHGTGTR  
TRIGDPLEAQHLLTGDRDARCPVWLGLSKNSIHAQAAAGVGIKMWLARRHGV  
LPTLHVDPSHVDMSAGGVRLTTEAREWPGVDRPRAAVSAFVSGTNAHLILEAP  
EALAEATDAPAPAPAPVDTVTEALEADATEAECAKAPGSPERAPAGVVP  
VVVSGRSVVVREAGRLAEVREAGVGLADVAVTMAGRSRFGTRAVLARGELAG  
RLRLAGAGDDAGVTVGAVVDPTGSGGGVVLVFPQGTQWVMGAGLGSSEVFAA  
SMRECARLSVHVWDLLEVVSGAGLERVDVQVPTWAVMSIARYWQMGVDVAV  
VGHSGQIEIAATVAGALSDEAAVVALRAGLIRGLAGGAAVPLPAGVEVREAGL  
KWPQVQVAVNGPASTVSGDRRAVAGYVAVCAQEGVQARLIIPDYASHRSHVEDLKE  
ELERLVSIGTRPRSPVCSVAGEQSPGFVQAGYFNRRLNRVFSVAVGGLLEGG  
HRRFIEVSAHPVLVHAIETAEAADESGEPFDAGYFNRRLNRVFSVAVGGLLEGG  
TWDPALPPGHLLTTPYFNNHHWAVTSPAGVGDAAAGFRGTMWEDHPFLRGLPLA

```

DSGRVFAGRLAGSHDLTDHVASGVTLTPTGTAFFVFAHAGAAATCGRLLELSVEA
PLVLPAGGVQVRVMSAADSSEGRVVAIHSAPAAVHSAEGDSAGVWTRUGEGTLL
VPDEPTPPADNARAMPAGERVEPAELVERFGALCYEYGEAPAGVRAWROPDALL
AEVLLPRASTGAGRGVHTFALLDALQPMWAGLLSVPEADAVLLPFAWQGVSLYATG
AGALVRLTKAGDGVSLQADTSGAAVLSLGAALVMPRLARRKLDVLLGSDAGERSLY
RVWQVRLPAGPSPRSWAVLGPDDADRLAGTFLGQDPDGGPTALYPFVRALRKALAAV
APREAVVTLVLSGAGATPESVOTTERCLTALQDMLDABELVDTPLVILTRGAAV
PGEIIGLACAGVWGLVRSRHPGFALVDTDGHDDRTALPLARVLDLGAQOLS
LRAGTAPVLLRAGTPEEROGPAPGATVLTAGTGLRLIARHLAAEHGVHLL
LLSGRAAGADLAELAGLEBPCEPAACDAADREALARYLAEVPADRLPTGVIIHA
AGVDDTGLDALPERITGVNRPKADALNLHEITRSPUSVFAVSGAAGILGRPGQ
ANYAANTFDLALAHRAHEDPLLVPMRLDLAALVREAEHGFDAVPGGLLGLPARAAV
THDCGLAFORALAAESADPLLVPMRLDLAALVREAEHGFDAVPGGLLGLPARAAV
ROAAVPYGGAPAGAGEGTAERLAGIAGEARLRELRLVRAEYGVILVSGPDVPEP
GRPKDGFOSLTAVELNRILRLPTALVDRPTSOAVAEYLAELAEIAGPRDGG
DTAAAEFEGDLAANAAGALABDDLRDLRLRLTELTAALTPQGRNPSAPAPASDL
DERLDSANDDLDFAFIEQL"
/misc_feature
/gene="tylg"
/function="encodes starter module"
942..2156
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase KS-Q of
starter module"
2571..3557
/gene="tylg"
/function="encodes acyltransferase domain of starter
module"
3675..3929
/gene="tylg"
/function="encodes acyl carrier protein domain of starter
module"
3993..8471
/gene="tylg"
/function="encodes module 1"
3993..5264
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 1"
5631..6617
/gene="tylg"
/function="encodes acyltransferase domain of module 1"
7410..7949
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 1"
9220..8471
/gene="tylg"
/function="encodes acyl carrier protein domain of module
1"
8541..13970
/gene="tylg"
/function="encodes module 2"
9541..9812
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 2"
10260..11246
/gene="tylg"
/function="encodes acyltransferase domain of module 2"
11319..11876
/gene="tylg"
/function="encodes dehydratase domain of module 2"
12861..13415
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 2"
13719..13970
/gene="tylg"
/function="encodes acyl carrier protein domain of module
2"
14351..19945
/gene="tylg"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"

```

```

/protein_id="AAB66505.1"
/db_xref="GI:2317861"
/translat="MTAELVATRKRLGALEREAREPIAVVAMSCRYPGVTPBDDLMR
LLADERDAVSLDRDGDLDALYDDPGPGTGYAREGGFLSHCHGCAFDALFFGLSPRE
ALAMDPOORLLTSTSEALERRAGVTDADRABSGRTGVYGVYDDYDARVLYGAGGPP
EDLEGYLVNSAGSIASGRVSYTFLGRLGPVNTVNTACSSIVSLHLAVRALNGECOM
REGRPLVAVIRGSAVNDQASNGLTAFNGRAQKQVIEDALDAGVQDQVDVAEHGT
GTBLGDPIEAGALLATYGTARTAEPLWGLSKSNIQHTQAAAGVAGVIMVLAARHG
RLPTTHVDPTTRTWKGVRLTTPVPVPGAEPRRAGVSSFGASGTNAHVLE
SVFAGEPAAGRPEDVTGGANTVSGRGAALRAQARLYDALTGTCGTGCGAGGAGP
GTAEVAGALAHARTAFRRHRAVVGCONRAELLAGLRELAEHEHPGPRVVTGTAPATER
TAFILFQGGQSRAGSGHLYRRHPVFAALDEVCALPEHLRPLRDLMPAESPRA
EPDRTEFTQFALQATLFLRAEHGRLAEALCGHSVGEIAAAHAGVLTLPDAAR
LVAAEGLMOLPAGGAMAALRAEAEIAPLERRAGELAAAVNGSVSSVVSDEAA
VLELLEOWRAEGREARLVASHAFSPHSDMDMLTQFDRVARTLTFAPTTLPVSLTGLG
TPVTEETLCTADHWVQAREPVLDMWTLRADGIDTFVELGPDGVLVSMARDCAAD
RPGDITTAGDGETPDPILLTLPRLRSVPETGDAEHGGFERALATAYAHGVPLRLAP
APDAASLVAALPTAFQTHWYLDAPAALPAGDDAGHPLLSAALDLPQGRGT
VWTGALSAATLPWAADSHVHSGRTVLPGLTALLDLALHAAPRGELTFEAPVLPEDGEV
RLRVLAEPDASGVRELSVHSGEDGGTWRHATAVLTGTTAGEPAGAPAAWPPGD
AEPDLAAEYERFADAGIGYGPAPRGLRSARWGDGAILADVRLPGLAGADRYGIHP
ALIDALHTAAALGGAGHMLPTTWNGVTLHARGAHAIIRVRLTPAGDPAVAVTDPA
LPGETPHPLASALADTGPAPRVLDLRIYGTATPRELASQALMALVQWQWADAEAA
EGRLVLTVRRAVDIGDGTVPAAATVWGLVRAAQSEHPGCFALLDDSPRSQULLPR
VAGTAEOLALRDGTLTLPASLTATLPAGARLPALDGTVLITGTCGSLGAEARHLVTR
HGARELLTSSRGQAPGAELVAELAAAGHADVAACDVADRAALALLRDLRVPAGHP
LTAVLHTAGVLDGCVLTQFQRLAALVLPKADAVRNHLHETTCGHASLAFILYSSAAG
VLGSAQSGVYAAANAYLDSFAVWRSRGLPAVSLGWGPDGGSAGSLGTTDLRLRR
Query Match 7.2% Score 36; DB 1; Length 43280;
Best Local Similarity 100.0%; Pred. No. 2.8e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 133 CTGGCCCCCGACGGCGCTGCAAGCGTTCGCCGAC 168
| | | | | | | | | | | | | | | | | | | | | |
Db 15074 CTGGCCCCCGACGGCGCTGCAAGCGTTCGCCGAC 15109

RESULT 10
SNA278573/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SNA278573 84985 bp DNA linear BCT 07-FEB-2003
Streptomyces natalensis pimarin biosynthetic gene cluster.
AJ278573.1 GI:12055067
ABC transporter; cholesterol oxidase; cytochrome P450 monooxygenase;
ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
dehydratase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
polyketide synthase; sensory transduction protein; thioesterase.
Streptomyces natalensis
Streptomyces natalensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyceinae; Streptomycetaceae; Streptomyces.
1
REFERENCE
AUTHORS
TITLE
Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N. and Martin, J.F.
A complex multienzyme system encoded by five polyketide synthase
genes is involved in the biosynthesis of the 26-membered polyene
macrolide pimarin in Streptomyces natalensis
Chem. Biol. 7 (11), 895-905 (2000)
JOURNAL
MEDLINE
PUBMED
20547809
11094342
2 (bases 1 to 84985)
Aparicio, J.F.
Direct Submission
Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
SPAIN
LOCATION/Qualifiers
1. .84985
/organism="Streptomyces natalensis"
FEATURES
source

```



```

/mol_type="genomic DNA"
/db_xref="taxon:68242"
<!-- 2086
/feature="ORF"
/feature_start=2
/feature_end=11
/feature_note="hypothetical protein"
/feature_id="CAC20917.1"
/db_xref="GI:12055068"
/db_xref="SPTREMBL:Q9EWA5"
/translation="TRPDAFREATPEYGYALLAPLMPERSADRVGKWTSDWSQARFL
ADHAYCEVLLTAGHGVLLLEDDQWADTASDLRLSPRCQGHPLGIVLTARKEE
TESDARTLRIVSDVGRPTATRLCGLPQAVGLVEAHAGSGVDPEVVEVHRSE
GNPYVMOLLSVQDARRLRDPAVDALLAHITPTGVREALNORFAALPEPVLRLRLC
AVGAEDVDTLLRRRTATCDASVAGLESARAGLGEDPHHPQLHFAHALVQETLFD
ELPRADRLRLKADALVARRHGVGDEIEIRVAHSHWAKSLPAARVLRLLLAA
EHAQMAEQVEIWLRRARHLVGLFPDDPATVRLLEQKHLQLGQVLTATRGYGHPE
AETALTRGALSTVTHAPEDPSPVLMALCAASLVTRYDASRQLGLRLDLSWQTRFPV
AVLGAAYGEGIVLHVGRMLPQALAELEHGVREADHFANRGHALARTFQHDPRVCSRY
NPTFWLHGHOIADARRELLSLTEYRSPDRSPALYVDGVVAAWEGDRTALASS
DSGVVAGDHGLLYKAMGLGVKGLVHAGDRDEGLALMHTSLAELGQSTVLRHPL
HGLGLQAQHQNGIQIEARTTFRMLLAAVEQRRERYVYLHPALPATLILHELLHGHGAE
ALADRDPRGRGV"
2190..3566
/feature="pimK"
/feature_id="CAC20917.1"
/db_xref="GI:12055068"
/db_xref="SPTREMBL:Q9EWA4"
/translation="MESARPIILFVPSGLLNLPLVLAGELSRQGVEDLWFATDEP
RNDVRIAEQSPVEFASGVDEMSAVTMSDEVREVTQPSRFAHRAVVRHTYRP
GLAEKFRRLQAVIDVQALMVIDICISGFANDAAIARNIYVYVPLFPVSNVLTAT
HPAKSVTPRGVPPVTLGRMTLQORVANVELPKLRTFAMFPLNPLKGLANDNRN
ELGLKASPMWPIEHADLVLCNSLELDPDIPEKMLVGMVPLPKAPDQDLSR
WLDQAQSVTVGLGTTTLRLTREQVGMVEVARLEDRHQVILWKLPSQQLHLIPRESL
PGLNRVESWVPMQMDLHAPHVQVFFTHGGGNGFNEGYPFGKPLVVRPLWVDCYQAV
RGQDGLSLTLDLPQTIDVNDVVDKLTVLGTGTPSYEKAERRAALRSAGGRETAAGL
VLSLPLA"
/feature="pimS4"
/feature_id="CAC20919.1"
/db_xref="GI:12055070"
/db_xref="SPTREMBL:Q9EWA3"
/translation="MENDKNDKIVDYLKXVNDLHQTQRRELVESAKOEPATIVG
MSCRLPGCVRSPEFWQLLKXDRGDIPTFDRGWDLDLAGGSGSSATAEGGFDIA
ASFDAGFGLSPREAVTMDPQRIILLESSWEALERAGIDVLSRGITSTGVFGTSGID
YINTVMSNRDIEGHSTGLIASILSGVSTYTFLEGPAVTVDTACSSLSVLHLAAQ
ALRVGECSLAGAGVIAISTPMSFVGFSGQGLAPDCKAFADTDGTAAAGSGVYL
VWRISIDARNHGVPLAVRGSANQDQAGNLTAPSGPQSVIRQALAAAGLSAD
VDVVEAGTGTALDPTLQALALTCYDRDRPRLILGSKNLGHTQAAAGAGLI
KMLAMRHGVLPRTLHVEARSSHVDSGALOLLTEPQEPKADRPWRAGVSSPISG
TNAILLEQPPAEEPEDEGPHTSAPVLPVLSAKSEERALTQGLDRLISALDRSPID
LGLSLATTRALHRAVLLSEGGEVVEVARGHAETAGKLVLPFGGQGAQRVNGREL
YGRFPFAEALDGLVIAHFDAGLDVMEGEAEGDDETGTQFALFAIEVALFRLESIG
VRDPFVAGHSITAAAHVAVGVLADACALVAARLMQALPAGGAMVAVQAESV
AARLDVGSITAAVNGPAAVITAGREERVLIAEFTGLGRKTRLPVSHAFSPLEP
MLEDPRVAGLSYEAPRIALVSNLTGELATDLVRAEYVVRHVRVTRFADGVSL
AAGASVFLGDPGVLTAMAQHTLDGTATVTPALRKDRSEETALLTALLAQHVGAGV
VDWMSGIFAGTGRVDLPYTPQFHRRYPRPVLSDGVGVAGLSRSHPLGLATLILA
ESGEVLTGRLSLQAPVPLDHTMTGGSVLPATGFLBLAMRAADQVGCDDRVDFVLT

```

```

PVLVAESGTQVQVLVGPADAGRTTISIHAHPDGDPOPMQHQAGTILTSERIVDF
GSTWMPQCAEAVDLEGFYDTTGYGPAFGQLRSVMQCEDGAYVEVLPQVADAGAF
GIHPVLLDAVLQSRMAGIGAGDTTFLPFAMRGSVSLHAGGASVLRARVKTGEDTLGL
VAVDAEGAPVLSVEBLVRLTQLTHTDPLTARRGAERDALLSLEWVSPEVPPAGVRC
VLSGADELGVASVATLADLSGEDFVVVPVSGVDVFAAAHETARTLGLLQEWSSG
REGTSASRLVYTRNADGGSVDVVAAMVGLVRSQAQSPGRIALVLEADSG
DLAEVLGLPGLLAAGDAQFVVREGAVRGRVLESNEGAGVSPREWDADGTVLVTGGT
AGVAVGVADVAHAHPTAVITHTAGVLDGDIASLTPERSAVLRPKVDAANLHEATKD
CGLGRELARHLVVERGVRLHLLVSGPEASGVGELRDLAAHAGIAVRAACDADRA
LTOALPYSYSSIGVLSGQGNVAGNVFLDALAGHAGLGLPAQSLAWGAWTQGGG
MUTATSDADMORMASVGTAPLSPERGLALFDAATAGSATHLVAVGVSVPVRAQGPV
SMRLGVKGRARTAVNTTGETAVSLADRLRAARPERPTOLMTDVVTEAAVLAADLAA
QNDVARFEYELGFSLTSTVEUNRLSTVIGRLSATVYVDSKNPTDLAARLYADLAA
QSAIDQASQAGFLVPAAPETLSDRMFLDALDSKVPEAQRMLSALGALPPEFNTA
ELEDLPLPATLAEGGTPELRI CVSTPANGVHEVYARLAASPRGERHVSALPLVGFAT
GERLPATAVTAARASADGNFVLVGHSSAGAFYLAALLENLTWIRPEAV
VILDTLSLHEQONESIDYAGLMRHEMVDVSPVMTNRSRLSAMASWGMGLNOLVVRH
TVTVLILPATKETFGITDGTGICYEDHSGSPVDVRSVDADHDFSMVRDDAPETARIVE
WLDLSGNA"
/feature="pimS3"
/feature_id="CAC20920.1"
/db_xref="GI:12055071"
/db_xref="GOA:Q9EWA2"
/translation="MTTSETVVAALRASLKEAENLRQNRKLVAAATEPIALVAMAC
RYPGVQSPEDLWELLTAGDAIGAFPTDRGWDLDALRDAGVDRRGNVQEGFLDG
VGDFDADFFGISPRESTMDPQRIILLESSWEALERAGIDVLSRGITSTGVFGTNGQ
DYAVLLVSLADATGIDGTIAASATSGRLSYTLGEGPAVTVDTACSSLSVLAHTAS
HALRAGECSIALGVGNVMSVSGLSMEFSRQGLAADRCKAFSDTADGTGWSGEGV
LVVERSDARNHGVPLAVRGSANQDQAGNLTAPSGPQSVIRQALAAAGL
DVTVAHGTGTPGLDPIEQAALLYTDQDRPERPILGSKNLGHTQAAAGL
KMLAMRHGVLPKTLIADVPSSHVDWTSGEVLLTEPQEPKADRPWRAGVSSFGLS
ITNNAHVIRQAEDEEPVPAEPVPLVSAKTEEAALQALARISSRGTASPV
VGFSLASGRSAFEHRAVLLAQCPDGLSEAAAGRTSRPLVLFSGQARQVGMGR
LVYRFPVFAELDEVLAHFDAGLRLVYMGDAEGLDGTGTPALFALEVALFRLAESL
GVYRDPVSGHSEIGETIAAHAVAGVSLTDACALVAARLMQALPAGGAMVAVQASEE
VVARLVDSYIAAVNGPEAVVLAGDETAVLRIAEELAAAGRTQRLPVSHPSPIME
PMLEDPRVRAEGLSGAPQIPLVSNVTGELAVEELVCSDPYVVRHVRVTRFADGVA
LEAGASAFLELGGDPGLTAMARHTLDETQVAVPLRKDRPEETALLTALAHLVAGI
GDWMSGIFAGTGARRVDLPYTPQFHRRYPRPVLVSGDVAGLRSAAHPLGLATLIL
AESGEVLTGRLSLQTHPLTDLHAI GNVFPGTGLELAHAAADQAGCDRVEDLVLM
TPIVLAEBNGTQVQVLVGPADAGRTTISIHAHPDGDPOPMQHQAGTILTSERVVD
FGSTWMPQCAEAVDLEGFYDTTGYGPAFGQLRSVMQCEDGAYVEVLPQVADAGAF
FGLHPALLDAVLQSHRIAGVSADDTQFLPFAMRGSVSLHAGGASVLRVVRVKTGDSMS
IAAVDAEGAPVLSAEAMVRAVYRDPGTQGAARHGAERDELLSLEWVSPEVPPAGVR
CVSLGADELGVASVATLADLSGEDFVVLPVSGVDVFAAAHETARTLGLLQEWSSG
GREGSGSRLVYTRNADGGSVDVVAAMVGLVRSQAQSPGRIALVLEADSG
GDIAELAGLPLGLLAAGDAQFVVREGAVRGRVLESNEGAGVSPREWDADGTVLVTGG
TGGLGRELARHLVVERGVRLHLLVSGPEASGVGELRDLAAHAGIAVRAACDADRA
AAVHALVAAPAEHPTAVITHTAGVLDGDIASLTPERSAVLRPKVDAANLHEATKD
DLILAFLVLSYSSIGVLPQANYAGNVFLDALAHRMGQGLPQASLAWGAWTQGG
GMTATLSDTDMORMASVGTAPLSPERGLALFDAATAGSATHLVAVGVSVPVRAQGPV
PSVIRGVKGRARTAVNTTGETAVSLADRLRAARPERPTOLMTDVVTEAAVLAADLAA
AKNVDARRENFELGFSLTSTVEUNRLSALTQGLSAALVDFDTPNPTALADHFPVSELY
EBAATGALLAELERLDSALGADPAPARTAAVTAARLAQIILDDWRDAPABEDAGVAE
RLEAASDEIFAFIDNELGRHSDR"
/feature="pimS2"
/feature_id="CAC20921.1"
/db_xref="GI:12055072"
/translation="MENDKNDKIVDYLKXVNDLHQTQRRELVESAKOEPATIVG
MSCRLPGCVRSPEFWQLLKXDRGDIPTFDRGWDLDLAGGSGSSATAEGGFDIA
ASFDAGFGLSPREAVTMDPQRIILLESSWEALERAGIDVLSRGITSTGVFGTSGID
YINTVMSNRDIEGHSTGLIASILSGVSTYTFLEGPAVTVDTACSSLSVLHLAAQ
ALRVGECSLAGAGVIAISTPMSFVGFSGQGLAPDCKAFADTDGTAAAGSGVYL
VWRISIDARNHGVPLAVRGSANQDQAGNLTAPSGPQSVIRQALAAAGLSAD
VDVVEAGTGTALDPTLQALALTCYDRDRPRLILGSKNLGHTQAAAGAGLI
KMLAMRHGVLPRTLHVEARSSHVDSGALOLLTEPQEPKADRPWRAGVSSPISG
TNAILLEQPPAEEPEDEGPHTSAPVLPVLSAKSEERALTQGLDRLISALDRSPID
LGLSLATTRALHRAVLLSEGGEVVEVARGHAETAGKLVLPFGGQGAQRVNGREL
YGRFPFAEALDGLVIAHFDAGLDVMEGEAEGDDETGTQFALFAIEVALFRLESIG
VRDPFVAGHSITAAAHVAVGVLADACALVAARLMQALPAGGAMVAVQAESV
AARLDVGSITAAVNGPAAVITAGREERVLIAEFTGLGRKTRLPVSHAFSPLEP
MLEDPRVAGLSYEAPRIALVSNLTGELATDLVRAEYVVRHVRVTRFADGVSL
AAGASVFLGDPGVLTAMAQHTLDGTATVTPALRKDRSEETALLTALLAQHVGAGV
VDWMSGIFAGTGRVDLPYTPQFHRRYPRPVLSDGVGVAGLSRSHPLGLATLILA
ESGEVLTGRLSLQAPVPLDHTMTGGSVLPATGFLBLAMRAADQVGCDDRVDFVLT

```

```
/db_xref="COA-Q9EWA1"
/db_xref="SPTREMBL:Q9EWA1"
/translation="MDQEQKURDYKLRASDLRRSRQVRGELERAAAEPTAIVCMSCR
PPGVRGPEELWLLAAGDGGITTSPPADRGWELLADAGEFSGGFLHDAEFDFGFFG
ISPREALAMDQOQLLLESSWEAFERAGIDPAAVRGTRTGMFGVMAQPOEYRVGPDVV
QGFALTGTTTTFVFNROGGI.SADGRCSFADSADGTGWSGVGLVLERSEARRNG
AGVTWMSPTTFVFNROGGI.SADGRCSFADSADGTGWSGVGLVLERSEARRNG
HEILAVKGSVAVQDASNGLTAPNGPSQSRVIBQALYSARLSADEVDVRAHGTGTT
LGDPEVAQALLATYGGDRDDRLPLLGKNSLHTQAAAGMAGVIXMVAHMGVLPL
RTLHVDPFSDHVDVTPQAVRLTEHPWPGQEQFRAGVSFGSGTGNATHLEAPPA
APSPAGGEDTAPVAEPTAAGAPVWLVSRTRREALRAQARLLEHLTSRPGIAPDVVA
YSLATARGLEHRAFTAAAGPAGARALATADGTPAPGLVQDARTARTSKLAFIPAGQ
GSORPCMGRELAARFPVPATLDELVGHFDEGLERPLKOVLFPAAREGTPAALLDQTY
AQPALPALEVALYRLAESFGITPDPFLAGHSIGEIAAHVAGVFTLADAAALVARGRL
MOALPEGAMVSLAEIEBEVLLLEDREDRSLIAVNGPAAVVIAGEADDVTAVEHF
DALGERTKRLRHSHAFSPLEMPLEAFERTAVSRUTPQAPVPPVSGLTGTLATVEQL
TSPDYWLHARHAFADALRLARRRHLPRMGPDGVLISAMAQSCCRGHVRRRY
LKHARRVPRPEAGLGAHHPLLTAASVAVANGDGLLTGRLSRRTHPWLADHVRGT
PKSLPATGDDVPAAGLGAHHPLLTAASVAVANGDGLLTGRLSRRTHPWLADHVRGT
VLLPGTAELELAVRAGDGCRCRVEELTAAPLVLPBEGGVQOVWVGPSPDASGRRAV
SVHSRDPGPEELPQTHAGTLVAGEHNGFATWNPADATFLDLDGFYARMATGFE
GYGPLEFQGLRAARWAGSDDDYAEVLPDSDGATSSSAFGLHPALDORLALHAAPVD
LGEDARGGLPFSQWDTLHASGASTVRVKLTPTDGDVAVALAVADTTGAPVASIGSLVL
Query Match 7.2%; Score 36; DB 1; Length 84985;
Best Local Similarity 100.0%; Pred. No. 2.6e-06; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 CGGCACGCGGGGTGCGCCCGACGCGCGCTGCAAG 156
Db 75833 CGGCACGCGGGGTGCGCCCGACGCGCGCTGCAAG 75798
RESULT 11
LOCUS AX250263/c 30000 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 3 from Patent WO0168867.
ACCESSION AX250263
VERSION AX250263.1 GI:15984065
KEYWORDS
SOURCE Streptomyces cinnamomensis
ORGANISM Streptomyces cinnamomensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Leadlay,P.F., Staunton,J. and Ollivnyk,M.C.
TITLE Polyketides and their synthesis
JOURNAL Patent: WO 0168867-A 3 20-SEP-2001;
Biotica Technology Limited (GB)
FEATURES
source
1. .30000
/organism="Streptomyces cinnamomensis"
/mol_type="unassigned DNA"
/db_xref="taxon:1900"
/note="Nucleotides 60001 - 90000 of the monensin
biosynthetic gene cluster"
ORIGIN
Query Match 6.6%; Score 33; DB 6; Length 30000;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 381 GGACGTCACGCGGTGGAGGGCGACGCGCACCGG 413
Db 23066 GGACGTCACGCGGTGGAGGGCGACGCGCACCGG 23034
RESULT 12
LOCUS AF440781/c 103450 bp DNA linear BCT 29-AUG-2003
DEFINITION Streptomyces cinnamomensis polyether antibiotic monensin
biosynthesis gene cluster, partial sequence.
ACCESSION AF440781
```

```
VERSION AF440781.1 GI:29122977
KEYWORDS Streptomyces cinnamomensis
SOURCE Streptomyces cinnamomensis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Ollivnyk,M.
TITLE The gene cluster for monensin biosynthesis
JOURNAL Thesis (1999) University of Cambridge
AUTHORS Ollivnyk,M., Stark,C.B.W., Bhatt,A., Jones,M.A., Hughes-Thomas,Z.A.,
Wilkinson,C., Ollivnyk,Z., Demychuk,Y., Staunton,J. and
Leadlay,P.F.
TITLE Analysis of the biosynthetic gene cluster for the polyether
antibiotic monensin in Streptomyces cinnamomensis and evidence for
the role of monB and monC genes in oxidative cyclization
JOURNAL Mol. Microbiol. 49 (5), 1179-1190 (2003)
PUBMED 12940979
REFERENCE 3
AUTHORS Ollivnyk,M., Ollivnyk,Z.V. and Leadlay,P.F.
TITLE The gene cluster for monensin biosynthesis
JOURNAL Unpublished
AUTHORS Ollivnyk,M., Ollivnyk,Z.V. and Leadlay,P.F.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2001) Biochemistry, University of Cambridge, 80
Tennis Court Road, Cambridge CB2 1GA, UK
FEATURES
source
1. .103450
/organism="Streptomyces cinnamomensis"
/mol_type="genomic DNA"
/strain="ATCC15413"
/db_xref="ATCC:15413"
/db_xref="taxon:1900"
/complement(<1. .1038)
/gene="gdhA"
/complement(<1. .1038)
/note="similar to glutamate dehydrogenase"
/codon_start=1
/transl_table=11
/product="glutamate dehydrogenase"
/protein_id="AAO65782.1"
/db_xref="GI:29122978"
/translation="MTTRPDTKALSQKTSLSQLTEIHRNPAQEFHQAAREVLET
LAPVIAARPEYAEAGLIERLCEPERQIVFVPMQDDHGRVVRNRFVFNESALGPYK
CGLRFPNSVNLGVIKLGPEQIENALTGIGGKGSGSDPRGSDAEVMAFCQSF
MTELYRHIGHDTVPAGDVGCGREIGYLGQYERITNRWEAGVLTGKRNWGSILIR
PEATGYGNVLFAMAALRRERGETLEGRTAVVVGSGNVALYTIQKLAALGANAVTCSDS
GYVDEKGDIDLKQVKEVARVDTYQRRGASARFVFPGRVNEVFADIALPFSATQ
NELDADATALI"
/complement(1220. .2140)
/gene="dapA"
/complement(1220. .2140)
/note="dapA"
/note="similar to dihydro-dipicolinate synthase"
/codon_start=1
/transl_table=11
/product="dihydro-dipicolinate synthase"
/protein_id="AAO65783.1"
/db_xref="GI:29122979"
/translation="MTLASLEPTPEFLNGLYVPLVPTFDDDLRLALEARLADAEA
LSAGASGLVALGTAAEATLTABERETVIRVCSAACRAHAGPLIVGTNDTATAITA
LRELAARGDVAALVAPPYIRPGEAGTLAHFAALAEHGLPLVVDIPLYRTGTLTGA
GTITAGRLPEVVGIKHATGSDIPTTWELDSPLPGFAVLGGDDIVLSPVAAAHGG
IVASANDRTADYAEIMIALWRGSAAPARLADIALRLSALFTEPNTVIKGLVLAQN
RIPSPAVRMPLLAASADSVRRAPLAASRK"
2211. .3152
/note="SCN 3"
/codon_start=1
/transl_table=11
/product="hypothetical DNA binding protein"
```

```

/protein_id="AAO65784.1"
/db_xref="GI:29122980"
/translation="MLDVGRLLHRLDELDRRTTIAVAEALFTTASVSQOLGVLREBA
GVPIILSGRRVVLTPAGSLVAHADLVNLRLEQAVAEIAGARDGIGGLIRLITGFPSG
GHIVPGALAEASHHPALEPMVRIDTSARVSDGLRAGELDLVALVDHYDFVPATPDTT
VDEPLPEEMFLYVTHAADTADTDSSGSTLAALLGPCASVPMITARDGTIGHAMAVRA
COAAGFPQRIHQVNDFRFVLDVALVAAGQGAGVFWRMAAESFGAGVVLTIKLPFRSKV
APRAGGGAHPALAAFAAATTATVERMAGSRGPAGGSE"
3264..3680
/note="SCN 4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO65785.1"
/db_xref="GI:29122981"
/translation="MADDAVLFLPDPRHLRGALAAGAVALGETCTETPAVHAWLQAHEA
SVSSEQRILPADAEILPKDAERLVPVISEERALKVEQCAPTVTDMESILLAFRE
TTQDQALVHRALTAGIPAQRIARLIGLDPFEIGRL"
complement(3684..4307)
/note="SCN 5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO65786.1"
/db_xref="GI:29122982"
/translation="MAVACAANVLPIDAVNRISAADVUVLFYFAYLLPYLAITMTWF
VSAVQVRSWARREARGTLQRYVLGTAFPGGSLSFIARAALVAVLWMLPHGLSTTF
SALPRTLVALALVVAAMICVVVAFVTFQADNLVENERALEFFGERSPANADYYFAL
AAMTTCGTTDDVTDSRMERTVAANTVIAEFVENTVTVALISALGGR"
4570..4758
/note="SCN 6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO65787.1"
/db_xref="GI:29122983"
/translation="MTVMDKLKQLMKGHEDKAGQIDKAGDFVDGKTQGYSGQVDTA
QDKLRDQFGSDQQBPPOR"
5058..5612
/note="SCN 7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO65788.1"
/db_xref="GI:29122984"
/translation="MGTAQSQEQAAPAACAAAFVRFVLGGVGVLASSFAVVALASWV
PWLANALVAVVTVVATHELHFPTFGAGRATWQHQAQSAGSAAAAYATCVAMEVL
QQLVACPAAVLEOVVYL SASALAGVARFVVLRLVVFARNRSLPAAAAVTRARPVRVP
APFATPVHASRPPAGPALCPAA"
complement(5693..6010)
/note="SCN 8"
/codon_start=1
/transl_table=11
/product="acyl carrier protein"
/protein_id="AAO65789.1"
/db_xref="GI:29122985"
/translation="WTSTDHTSQGDATELEKOLAATPBREKLLITDTIRTQAGTLIN
TTLSDDSNFLENGLNSLTALLETIKLTIITMGIEIAMVAIVENETPAQLAHHLGQELAH
TTA"
complement(6045..8531)
/note="SCN 9"
/codon_start=1
/transl_table=11
/product="ketosynthase-like protein"
/protein_id="AAO65790.1"
/db_xref="GI:29122986"
/translation="MANEEKLYVKWTTAEIHAQOOQLRELKAAOEPIAVVSMACK
LPKTKPDDLWDLVSEGSDAVTGFPDDRWEIPEERPYPALGGLFDADAAGFDAGFPD
ISDTEAVATEPLQRIAMLHAWETVERGHIAPHITLSTGYVYGATGDHYATRELETPA
DELLPYLGSGTSPHTFHAFHOKSLAQDCRKPFAAADGMGLGEVGLVLERIGCGGL
ALAGGQVMTSTPTHFAFKSLAQDCRKPFAAADGMGLGEVGLVLERIGCGGL
KNCHPVLATRVGAINVCARESDNEPLNGSVQHSNTGCTOGCAAGAILKNVOAFRHG
CNCISALATRVGINVCARESDNEPLNGSVQHSNTGCTOGCAAGAILKNVOAFRHG

```

CDS

gene  
complement(8643..9542)  
/gene="monCII"  
/complement(8643..9542)

CDS  
/gene="monCII"  
/note="monensin epoxide hydrolase; MonCII"  
/codon\_start=1  
/transl\_table=11  
/product="monensin cyclase"  
/protein\_id="AAO65791.1"  
/db\_xref="GI:29122987"  
/translation="MNKLRI PVSVQTVSLNVVRVPADPGAPGRPFLLHGLMSLNARMW  
DEVARIALAAGHPAYAVDRHGCHESDTPDGYNATVTDLVAAVTALDLSGALVAGH  
SWGHALRLAEAPDLPVLAGLIDGWYFDFGPGVMRAFETVVRRAQOQTTSAA  
DMRAYLRATHPDWSPTSIEARLADYRVGPDGLLI PRLTSTQVMSI VAGIQREAPADY  
PKVTVPVRLPLDIPALPQLSDQVRARWAABEALEQVSVRWYPGSDHDLHAGADEIA  
ADLLLARSCEAMPGGKAGVREA"  
complement(9596..10426)  
/gene="monE"  
/complement(9596..10426)  
/gene="monE"  
/note="SAM dependent; MonE"  
/codon\_start=1  
/transl\_table=11  
/product="monensin 3-O-methyl transferase"  
/protein\_id="AAO65792.1"  
/db\_xref="GI:29122988"  
/translation="MNKTVAPEPSDIGHYVDHKVFDLMTOIGDGNLHYGYWFDGEGEQ  
AFIDEAWQVOTDENIRRLDPAPGDRVLDICGGNGTPAMOLARARDVEVVGISVARQ  
EKGNRARAGLADRVFEQVDANLPLFDDGSDHCHWALSMLHMPKDQVLTAEHR  
VKPGRMPADMYVLPNPDRPRPTATVDTTYAALTDIGDYDIPDI FRAGTWVLETL  
ITRETAKTVDGYEWIRAHREDYVDIIGVEGYELFELHNQALGKMPELGYIFATAQR  
"

gene  
10656..12191  
/gene="mont"  
10656..12191  
/gene="mont"  
/note="membrane transporter; putative monensin  
transporter; Mont"  
/codon\_start=1  
/transl\_table=11  
/product="putative monensin resistance protein"  
/protein\_id="AAO65793.1"

Query Match 6.6%; Score 33; DB 1; Length 103450;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 381 GGACGTGCACGCCGTGAGGCGCACGCCGCGG 413  
|||||  
Db 83066 GGACGTGCACGCCGTGAGGCGCACGCCGCGG 83034  
|||||

RESULT 13  
SERERYAB S. erythraea second and third ORF's of eryA gene, complete cds.  
LOCUS 20235 bp DNA linear BCT 26-APR-199  
DEFINITION M63677  
ACCESSION M63677.1 GI:150693  
VERSION 6-Deoxyerythronolide B; acyl carrier protein; acyltransferase;  
KEYWORDS beta-ketoacyl-ACP synthase I; beta-ketoreductase; dehydratase;  
enoyl reductase; eryA gene.  
Saccharopolyspora erythraea  
Saccharopolyspora erythraea  
Actinobacteria; Actinobacteridae; Actinomycetales;  
Bacteria; Actinobacteria; Pseudonocardiaceae; Saccharopolyspora.

1 (bases 1 to 20235)  
Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. and Katz, L.  
Modular organization of genes required for complex polyketide  
biosynthesis  
Science 252 (5006), 675-679 (1991)  
91220065  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1. .20235  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1836"  
19. .20235  
/gene="eryA"  
19. .10722  
/gene="eryA"  
/note="eryA ORF2 encoding modules 3 & 4 for  
6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26494.1"  
/db\_xref="GI:152694"  
/translation="MTDSEKVAEYLRRATLDLRAARQIRELESDPIAIVSMACRLPG  
GVNTPQRLMELLRGEGTILSGFPDRCMDLRLHHPDNGTSTYVDKGGFLDDAAGF  
DAEFGVSPLEAAMDPOORLLETSMELVNEGIDPHSLGATGTFVLGAKGYGE  
DTAAAEVDGVSVTGAPAVASGRISYTMLEGPSISVDTACSSSLVAHLHVASLGRK  
GESSMAVVGAAVMTGCVFVDFQRALADGRSKAFAGADGSGFSEGVTLVLLER  
LSEARRNGHEVLAIVRGSGALNDGASLAPSGPAQRVIRQALESGLPEFGVDVAV  
BAHTGTALGDPLEIANALLDVTGDRDADRPLMLGSKVKSNIQHTQAAAGVTGLLKVVL  
ALRNGELPATLHVSEPTPHVDWSSGVALLAGNOPWRGERTRARVSARFISGNTAH  
VITTEAPEREHRETTAHDGRLVPLVVSARTTAALRAQAQAIAELLERDADLAGVGLG  
LATTRAREHAAVVAATREAEVRLREIAGAATAADAVVEGTEVDEGRNVVFLPQO  
GSOWAGMAELSSSPVAGKIRACDESMAPQMDKVSVDLRLQAPGLDKRDVVPQ  
VLEAVMSLAELWRSYGPVAPAAVVSQGEIAAHHVAGALTLEDAKLVGRSRLMS  
LSGGGMAAVALGRAVRLRPMQDRLSVAANGPVSWSGEPGALRAFSECCAAE  
GIRVDDIDVYASHSPQIERVERLETTGTDIAPRPVTFHSTVESRMDGTELDAR  
YWYRNRETVPADPRLAASGYDAFIEVSPHPVVOAVEAEVDEAGDADVVGGS  
LHRDGDLSAFRLGNATVHSGVDLRMDVALPAAFPALPTTFPORKWYLIQAPAA  
ASDELAVYFWTPIEKPSGNLDGDLVVTPLISPEWTEMLCEAINANGRRLCEVD  
TSARTEMAQVAQAGTGRGLVLSLSDGSAACPGVPAGVGLITLVQALGDAGDA  
PWCLTQGNVTPADDDLARPAQTTAHGFAOVAGLELPGRWGVVDLPESVDDAARL  
LVAVRGGRAEDHLAVDRGLHGRVVRASLPQSGRSWTGHTVLTGASPVGDQ  
LVRLADRAERLVLGACPGDGLLAARVEEAGASAVCQDAAALREALGDPVTVLV  
HAGTITNFGTSIEVAFPEAETIAAKTALLAVLDEVLGDAVEREVISSVAGIWGGA  
GMAAYAGASVYLDALAEHRRARGRCTSVAMTPALPGGAVDDGYLRERGLSLADR  
AMRTWERVLAAGPVAVADWVPLSEGFATRTPALFAELAGRGQAEPDPSGT  
GEPARLAGSPDEQOENLELVANAIEVGHESAAEINVRRAFSELGDSINAMAL  
RKRLSASTGLRLPASLIVDPHTVTALAOHLRLVGLDQADQAAVRVVGADESEPIAV  
GIGCRFPFGIGPEQWRLVLAGANLTTGFPADRGWDIGRLVHPDNPNGTSTYVDKGG  
FLTDADDFPGFGITPREALAMDPOORLLETSMELVNEGIDPHSLGATGTFVLG  
MNGQSYMQLLAGEAEVVDGYQGLNSASVLSGRIATYTFEGEPALTVDACSSLVGI  
HIAQALRGECSLALAGGVVMSDPTTFDFSTQGLASDGRCKAFSARQFALASE  
GYAALVLEPLSARANGHOVLAIVRGSAVNDGASGLAAPNGPSQERVIRQALASG  
VPAADVVEAHTGTGELGDTIAGALIAATYQDDRRLRLGSKVKNIGHTQAAAGAA  
GVTKVLAWRHMLPRLSLRADELSPHLDWESGAVEVLRVWPWAGERPRRAGVSFG  
VSGTNHVIIEEAPAEQEARTEARTEGRLPVLSGRSEAVVAAQARALAEHLRTPELGL  
TDAWTLATGRFADPRAVVLGDRAGVCAELDALBGRPSADAPVTSAPRKPVIIV  
FPGQAGQWGMARDLDESSEVFAESMSRCAEALSPHTDWKLLDVVRGDDGPPDHERV  
VLQPLVLSIMVLSAELWRAGVTPAAVGHSGQEIAAAHVAGALSLEAAKVVALLRSQ  
VLERLDDQGSVGVASHSPSEVARIEDRLAEGLTITAVRGSVPLSTVTGVEIDTSA  
AEAREMKPRRIAVYASHSPSEVARIEDRLAEGLTITAVRGSVPLSTVTGVEIDTSA  
MDASVYRNLRRLRVLFEQAVRGLVEQGTDFTEVSPHPVLLMAVEETAHAGAEVTCV  
PTLRREOSGPHFELRLRAHVHVGAGDLFPVAGRGPAELPTTFPHQRFPVPHRP  
ADVALAGVDEGLPVLIELVLRPLVLAGAGALLRMSVGAPDESGRRTIDVHAEDVA  
VDLAAQVSGEATGTLACVAGARDTEQWPPEDAVRILPDDHYDGLAEQGYGSPFO  
DLAQAQMSQRATGTLACVAGARDTEQWPPEDAVRILPDDHYDGLAEQGYGSPFO  
ALRAAMRKDSDSVVAESIADEEGIAFHVLLDVAQTLISLIGALGEPGGKLPAMNT  
VTLHAGSATSRRVATPAGADAMALRVTDPAHLVATVDSLVVRSTGEKWEQPEPRGG  
EGELHALDWLRAEPGSGTRVADASDLDAVLRSEGEPEFPAVLRYEPEGGDPPRAA  
RHGVLMAALVRRLEQELPGATLVIATSGAVTSDSDSVPFEGAAAMGWVIRCAQA  
ESDPFRVLLDLDAAEPGMLPAVPDPNPQLRGDDVFPVRLSPLASALTLPAGTQRLVP

GGCAIDSVAFEPADPVEQPLRAGEVRVVRATGTVNFRDVLALLCMYPQKXADMCTEAG  
VVTAVGPDVDAFAPGQVILGFCQAFAPITAVTDHRLIARVPDGNWDADAAPVIAVTT  
AHYHDLHLAGLRAGQSVLIHAAAGVGMAAVALRAGAEVLAATAGAKHGTIRALGHL  
DOHAISSRETGPAKFRERTGRGVVDVNLSTBELLDSEADLLAEADVFEVEMKTD  
LRDAGPRGRYAPFDLGEAGDDRLGRLREVUGLLGAGELDRLPVSAWELGSLVPAALQ  
HMSGRHVGLVLTQAPVDPDPTVLTITGCTGTLGRLLAHLVTEHGVRLHLLVSRRG  
ADAPGSDRLAEITDIGASAEIAACDTADRLDALSLDGLPRLTGVVHAAGVADGL  
VTSIDPAVEQVRLRAKVDAAWNLHETLANTGLSFVILFSSAASVLAGPGGQVYAAAE  
SLNALAALRRTRGLPAKALGWLWAQASMTSGIDRIARTGVAALPTERLALALFDSA  
LREGGEVVRPLSINRSALRAAEVPEVLKGMVRKALRAQGAARAGNVVDRLAGRSE  
SDQVAGLAEVLRSHAAVSGVGSADOLPERKAFKOLGFDSLAAVVELNRILGTAVRL  
RSTLVDHPTPLAVAEHLRDLFAASPAVDIGDRLDELEKALEALSADCHDDVQGRIL  
ESLLRRWNSRRADAPSTSAISEDASDELFSMLDQREGGSDI"  
19. .4470  
/gene="eryA"  
/note="putative"  
/function="approximate span of module 3"  
97. .1482  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
of module 3"  
1693. 2670  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 3"  
3406. 3921  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 3, possibly non-functional"  
4171. 4428  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 3"  
4471. .10722  
/gene="eryA"  
/note="putative"  
/function="approximate span of module 4"  
4471. 5847  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
domain of module 4"  
6054. 7026  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 4"  
7165. .9216  
/gene="eryA"  
/note="putative"  
/function="approximate span of dehydratase and  
enoylreductase domains of module 4"  
9433. .9984  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 4"  
10225. .10483  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 4"  
10723. .20235  
/gene="eryA"  
/note="ORF3 encoding modules 5 & 6 for  
6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26495.1"  
CDS

```

/db_xref="GI:152695"
/translation="MSGDNGMTEKRLRYLKVTVTSLDSTVRLREVEHRAEPIAIV
GMACRFPDVPSPFSEFVSGGDAIAEAPADRGWEDDPDARLGMLAAGDPDAGP
FGISPREALNDPQRIIMLEISWEALERAGHDVPVLSGATGCVFTGVTGVIYGRPD
APDEVLYGVTGTGASVAGVAYCGLGEPAMTVDTTSCSLGTALHLMASLRDEC
GLAAGLVTVMSGPAFTFPGSGLAADGRCKPFKAADGFGLAEGAGVILVQLISA
ARREGRPVLVRSAGVNDQASGLTAPSGPAQQRVIRRALENAGVRAGDVVYEAH
GTGRLGDPVIEHALLSYGABRPDDPLWIGSVKSNIGHTOAAAGVAGVMAVLAIR
HEMPRTTHFDPSPQIEMDLGANSVVSQASRWAPGRPRRAGVSSFGISGTVNAHIV
EEAPADPEFPDGPVPLVLSGRDEQMRQAGRLARGLADHLAREPNSRLRTGFTLAT
RRSAMVHRVAVVVDRLDALAGRAVDIADRTATGQARTTRGVAMVPPFGAOWQG
MARDLLRSOVPADSIKDCERALAPHVMSITDLISGARPLRDVDPVAFVMSV
AALWESHGVEPAEAAVAGHSQETAANAVAGALTLEDAKIVAVRSVLRLGLGQGMAS
FGLGTEQAERIGRFGAGLSIASVNSPRSVVAGSGPLDELIAEACEAERIGTARRIPV
DIASHSPQVESLRRELLTELAGISPVSDADVALYSTTGQPIDTAMDTAYVYANLEQ
VAFQDRLQAEAGDFAEVSPFVPLTVGIEATIDLSALPADAGACVVGTLREDRGL
ADFTHALGAYAQGVDEWSPAPADARVPELFPFQRYWLPITPGGRARDEDDDW
RYQVVRREAMESIAGVRLVLTGTPVSELSDAIRSGLEQSGATVLTCDVESRSTI
GTALAEADTALSITVLSILSRDGEAVDPISDALALVQALGAAGVAPLWVLRNAVQV
ADGELVDPQAQMVGGIGRVVGLIOPCRMGILVDLDADAASIRSLAAVLADPRGEQV
AIRADGIKVRLVPAAPARATMSPRGTIVITGTTGGI GAHVARWLABSGAEHLVL
GRGADAPASERBELTALGTGTITACDVADRAKLEAVLAAERABEGRTVNAVHAA
GVSTPTDLDTAEETETADYKVTGTVNLDELCPDLDAFVLFSSNAGVWGSFGLASY
AANAFLDGFARRRSEGAFTVSIAMGLWAGQNMAGEGYELRSGLEAMPDPRAVE
ELHITLDHGQTSVSVVDMRRFVELFTAAHRPLFDELAGARAEARQSEEGPALAQ
LAASTAERREHLAHLIRAEVAVLGHGDAAIDRDRAFDLIGDSMTADLNRNLA
VTGVRRAATVVDHPTITRLADHYLERLVGAEEQAAPALVREPKDADDPILAVGMA
CFPPGVHNPGLWELFVGRGDVAEMPTIDRGMDDLALDFDPDQRHGTYSYRHGAFLD
GAADFAAPFGISPREALNDPQRIIMLEISWEALERAGHDVPVLSGATGCVFTGVTG
QYGVADVPPEISGVLGTCNSAVSVGRVAYVGLGEPAMTVDTTSCSLVLAHSHAC
GSLRDGCVGLAVAGSVVAGPEVTFEFGKGLVLDGRCKAFSAADGFGAEVAGTGA
VILLORLSDARRAGROVLGVVAGSAINQDGSNGLAAPSGVAQQRVIRKAWARAGITGA
DVAVBAHGTRGLGDPVIEHALLSYGABRPDDPLWIGSVKSNIGHTOAAAGVAGV
IKVVLGNRLVPLMLCRGERSPLTSEWSSGVELAEAVSPWPAAQGVRRAGVSAFGV
SGTNAVITIAEPEPELPEPPEPGLVLAANSVPLLSARTETALAAQRLLESADD
SVPLTALASALATGRAHLPRAALLAGHDHQLRQGLRAVAGVAAFCATTGPTASAGV
VVFVPGQAGQWCMARGLLSVFVFAESIAECDVLSVAGVSAFSEVLEQRPDAPSLER
VYVQVPLESVYSLRLMAGCGVSPASVIGHSGQETAAVAVAGVLSLEDGVRVVALR
AKALRALAGKGVWLSAOPGERARALIAWEDRISVAVNSPSSVSVSGDPEALAEIV
ARCEBGVRAKTLFVYASHSRHVEIRETILADLDGISARRAAIPLVSYLHERRDG
ADMGPRVYVNLRSOVPADSIKDCERALAPHVMSITDLISGARPLRDVDPVAFVMSV
LHRDTABEHLIAELARAHVGVAVDMWRNVFPAAPVPLNPPYPPQRYWLAPEVSDQ
ADSRVYVDRWELATTPVLDLGGFLVHGSAPESLTSAVEKAGGRVVPVASADREASAL

```

```

Query Match 6.4%; Score 32; DB 1; Length 20235;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 413
DB 1006 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 1037

RESULT 14
LOCUS AR049368 20235 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5824513.
ACCESSION AR049368
VERSION AR049368.1 GI:6005407
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.
TITLE Recombinant DNA method for producing erythromycin analogs
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;
FEATURES
source
1. .20235
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

```

```

Query Match 6.4%; Score 32; DB 6; Length 20235;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 413
DB 1006 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 1037

RESULT 15
LOCUS AR095529 20235 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 3 from patent US 6004787.
ACCESSION AR095529
VERSION AR095529.1 GI:10023473
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.
TITLE Method of directing biosynthesis of specific polyketides
JOURNAL Patent: US 6004787-A 3 21-DEC-1999;
FEATURES
source
1. .20235
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 6.4%; Score 32; DB 6; Length 20235;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 413
DB 1006 GACGTCGACGGGTGGAGCGGCACGCGCACCGG 1037

RESULT 16
LOCUS SEERVABS 20444 bp DNA linear BCT 28-FEB-1992
DEFINITION S.erythraea eryA gene for 6-deoxyerythronolide B synthase II & III.
ACCESSION X62569
VERSION X62569.1 GI:46977
KEYWORDS 6-deoxyerythronolide B synthase; acyl carrier protein;
acyltransferase; beta-ketoacyl synthase; dehydratase; enoyl
reductase; ketoreductase; multifunctional subunits; thioesterase.
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.
REFERENCE 1
AUTHORS Bevitt, D.J., Cortes, J., Haydock, S.F. and Leadlay, P.F.
TITLE 6-deoxyerythronolide B from S. erythraea: cloning of the structural
gene, sequence analysis and inferred domain structure of the
multifunctional enzyme
JOURNAL Eur. J. Biochem.
REFERENCE 2 (bases 1 to 20444)
AUTHORS Bevitt, D.J.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-1991) D.J. Bevitt, Biochemistry Dept., Univ of
Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK
COMMENT For related sequences see X56107 & M63677.
FEATURES
source
1. .20444
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/strain="NRRL 2338"
/db_xref="taxon:1836"
/clone="SE55, DE10, pBK25"
/clone_lib="genomic in lambda EMBL3"
37. .42
46. .20268
RBS
gene

```

```
/gene="eryA"
46..10749
/genes="eryA"
/notes="ORF B"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase II"
/protein_id="CAA4448.1"
/db_xref="GI:581651"
/db_xref="GOA:Q03132"
/db_xref="SWISS-PROT:Q03132"
/transl_table="MTDSEKVAEYLRLRLDRAARQRIRESDDPIAIVSMACRLPG
GVNTPQRLWELLRGSTLSCFTDRCMDIARLHHPDNPDTGYVDKGGFLDAAAG
DAEFFGVSPREAAANDQOORLLLTSMELVENAGIDPHSLRGATGTVFLGKGYG
DTAAEDVEGVTGAPAVASGRISITMLEGFSISVDFTACSSSLVALHVLAVSLRK
GESSAVGVGAVNATPGVDFDQRALAADGRSKAPGARVIRIQALSCGLEPGBDVA
LSEARNGHVLAVRGSALNDCGASGLSAPGARVIRIQALSCGLEPGBDVA
EAGTGTALGDPLEANALLDYGDRDADRLMLGSKVNSNIHTQAAAGVTGLKVL
AIRNGELPATLHVEPTPHVDMSSGVALLAGNQPWRGRTRRAAVSAFGISGTNAH
VIVEAPREHRETTAHDGRVPYLVVSARSSTAALRAOAAQIAELLERPDADLAGVGLG
LATTARHEHRAAVASTREAVRGLREIAGAATADAVEGVTEVDNRNVLFPQG
GSQWAGMGAELSSSPFAGKIRACDESMAPQDWKVSVDLRQAPAGLGRVDVQP
VLEAVVLSAELMRSYGEVPEAAVGHSGQCEIAAHVAGALTLEDAKLVGKSRMRS
LSGEGMAVALGEAAVERLRPMQDRLSVAANGPRSVSGEPGALRAFSEDCAAE
GIRVDIDVYASHSPQTERVEELTGTDIAPRPVTHSTVSESMDCTELDAR
TYNRNIRETPADATLAEISGYDAFIEVSPHVVOAVERAEEADAEDAVVGS
LHRDGDLSAFLRSWATIRSDVLRWDVALPGAAPALPTYPFORKEYWLOPAPAA
ASDELAYRVSPTIEKPSGNLDGWLVTPLISPEWTEMLCEALNANGRALRCVEYD
TSASRTMAQAQAQAGTGFGRVGLSLSDESACRPGVAGVGLLTVALQADAGDA
PWCLTQGAARTPADDLARPAQTAAHGAQVAGLELPGRWGGVLDLPEVDDAALL
LVAVLRGGRAEDHLAVDRGLHGRVVRVAPLSQSSRSRTPHGTVLVTGAAPVGDQ
LVNLRADRAERLVLACPGDDLLAAVEEAGASAVCAQDAALREALGDRPVALV
HAGTUTNFGSIEVAPEEPAETIAKTAALLAVTDEFGDRAVEREYCSVAGIWGGA
AMRWERVLAAAGSAYDLAEHRAGRSCTSVAVTPMALPGGAVDDGVLRRGLRSLNDR
GPAORLACLGSDEQOENLELVANAVERLHESAAEINVRRAFGSELGLSLNVAL
RGLSASTGLRLPASIVRPHPTVTAALQHLRLVGDADQAAVVRVGADESEPTAIV
GIGCRPGGIGPEQLMRVLSEGNLTTCFPADRGWDIGRLVHPDNPDTGYVDKGG
FLTDAADFFGFTITPREALANDPQORLLMTAWEAVERAGIDPALAGTDTGVFG
MNGOSYMQILAGEAEADVGGYQIGNSAVLSFGRIATFGWEGALTDVTCASSSLVGI
HMAQALRGECSLALAGVTVMSPDYTFVDFQRGLASDGRCKAFPSQIRQALASG
GVAALVLEPLSRARANGHOVLAVLRGSANVQDGASGLAANGPSPQIRVQALASG
VPAADVVEAHTGTGTGLDIEAGALITATYQDORLRLGLSGVKTNIHTQAAAGAA
GVTKVLMRHGMPLASLHDELSPHLDWESGAVELREVPWAGERPERACVSGFG
VSTANTHATVIEEAPAEARTEKRGELFVLSGRSEAVVAQAARAEHLRDPBELG
TDAWTLATRAEDFVRAAVLGDGRAGVCAELDALAEGRPSADAVAPVTSAPKPVLY
FPOCGAWGMARDLLESSEVFASMSRCAEALSPHTDKLLDVVRGDGPDPERHVQ
VLQVLFISIVSLAEIWRHAGVTAAVGHSGQCEIAAHVAGALSLEAAKVVALLRSQ
VIRELDDQGMVSGARDELTVLARWDGRVAVAAVNGPSTVAGPTAELDEFFAE
AEREMKPRRIAVRYASHSPQTERVEELTGTDIAPRPVTHSTVSESMDCTELDAR
MDASVYTRNLRVPLFQAVRGLVEQDFTFVUSHPHVLMAVEETAEHAGARVTCV
PTLRREGQSPHEFLRLNLRHAGVGGADLRPAVAGRPAELPTYPFHEQWMPRPHRP
ADVALSGVGAEPHPLLLAAVDVPHGGAVFTGLRSTDEQPLWAEHVGGGTLVFGSVL
VDLAAQWSDVGLPVLBELVQLRPLVLGAGALLRMSVGAPEDESGRTIDVHAEDVA
DLADAQWQATGTLAGVAGRPTEQWPEDAVRIPLDDHYDGLAEQGYEYGPSP
ALRAMKDSVYAEVSIADDEGVAEPHVLDAVQTLISLGAELGEPGKGLPAMWT
VTLHASGATSVRVVATAGDAMALVTDPAHLVAQTLISLVTGCKEWEQPRPGG
EGEHLNLDVRLAEPPGSTRGAGVDDVNLSTGELLDESADLLAEDQVTEVSGKTD
DDEHATSRETFARKFERFGGRGVDDVNLSTGELLDESADLLAEDQVTEVSGKTD
LRDADFRGRVPALEQBELPGATLVATSGAVTSDDDSVPEGAAMWGVICQAQ
ESPRFLVLTDAEPGLMPLPVPNPQALRGDDVFPRLSPAPSAULTPATGRLOVP
GDGADSVAREPAPDVEQIRAGEVRVDRATGVNFRDLVILALGMYPQKMDGTEAAG
VTVAGPDVAPAGDVRVGLFQCAFAPVATDRLILARVDPGMSDADAAPVTAFTT
AHYALHDLIAGURAGSVLIIHAAAGCGVMAVALARRAGAVLATAGPAKHTLRAIGL
DDEHATSRETFARKFERFGGRGVDDVNLSTGELLDESADLLAEDQVTEVSGKTD
LRDADFRGRVPALEQBELPGATLVATSGAVTSDDDSVPEGAAMWGVICQAQ
HMSRGRVGLVLTQAPADPGTVLTIGTGTGLRLHARLHTEHGVHLLILVSRG
ADAPGSDBLAEIEDLGSAAETACDTRDALGALLDLPLPLTGVVHAAGVLADGL
VTSIDEPAEQVLRKADVAAWNLHETANTGLSPVLFPSSAASVLAGCGQVYAAANE
SLNALAALRRTRGLPAKAGLWGLWAQASEMSTGLGDRITARTGVAALPTERLALPDSA
LRRGEVYVPLISINRSLARRAEFVEVLRGMVRKALRAAQAEAGPNVDRKLARSE
SDOVAGLEHTRSHAAVSGYSDQLEPKAFKDLGFDLSLAELNRNLGTATGVRL
PSTLVDFHPTPLVAEHLRDLFAASPAVDIGDRLEDEKALEALSAEDGHDVQGRLL
ESLLRNRNSRADAPSTSAI SEDASDDDELFSMLDQRFEGGDEL"
```

```
139..1416
/genes="eryA"
/product="beta-ketoacyl synthase"
1471..2763
/genes="eryA"
/product="acyltransferase"
4261..4455
/genes="eryA"
/product="acyltransferase"
4513..5787
/genes="eryA"
/product="beta-ketoacyl synthase"
5839..7116
/genes="eryA"
/product="acyltransferase"
7117..7563
/genes="eryA"
/product="dehydratase"
8743..9234
/genes="eryA"
/product="enoylreductase"
9418..10164
/genes="eryA"
/product="ketoreductase"
10315..11059
/genes="eryA"
/product="acylcarrier protein"
10750..20268
/genes="eryA"
/notes="ORF A"
/codon_start=1
/transl_table=11
/product="6-deoxyerythronolide B synthase III"
/protein_id="CAA4449.1"
/db_xref="GI:46979"
/db_xref="GOA:Q03133"
/db_xref="SWISS-PROT:Q03133"
/transl_table="MSGDNGMTEELRRLVLTVELDSVTARLREHVEHAGBPIAIV
GNACFPDGVDPSPFSEFVSGGDAIEAPADRPDPDARLGMALAAAGDFDAGF
FGISPREALANDPQORIMLEISWEALREHDPVSLRGATGTVFGTVGVYGRPDE
ADEVLGYVTGTGTSVAGSRVAYCIGLEGPMTVDTACSSGLTALHLAGESLQRLSA
GLAGAGVYVMSPGATFRFSGQGLAADGRCKPFSKAADGFLAEGAGVLTQRLSA
ARREGPVLAVLAGSVNQDGSNGLTAPSGAQRVIRRALENAGVAGDVYVEAH
GTGTRGDPLEIHALISTYGAERDDPDLNIGSVKSNIGHTQAAAGVAGVAKVALAR
HEMPRTLHDFEPSPQIEWDLGAVSVVQSARSWPAGERPRRAGVSFEGISGTHAVTV
BEAPADEPEPAPDSGFVPLVLSGRDEQARQAQGLADHLEPEPRNSLDTGFTLAT
RASAMHRAVVGDRDEALAGRAVADRIRATATCGGNSPRRVAVVPPGQGAQW
GMARDLRSQVFPADSRDCERALAPHVDSITDLSGARPLDRVDVQRLGQGGMA
LAALWRSHVPEPAVVGHSQGTAAAHVAGALTLEDAALVAVRSVRLRLGQGGMA
SFLGTGEQAARIGRFPAGALSIAVNGSPSVVVGESGLDELIAECEAEAKARRI
PDYASHSPQTERVEELTGTDIAPRPVTHSTVSESMDCTELDAR
EQVRFODATQALAEAGDFAFVESPSPHPLDVTGTEATLDSALPADAGACVYGLLRBG
GLADFTALGEAYAQGVVDWSPADAREVELEVPYFQRYWLPITPGGRARDEDDM
RVOVWREAESASLAGRVLLVTPGVPSELSDAIRSGLEQSGATVLTCDVBSRSTI
CTALEADTALSTVGVAAVPHGEAVPSLDLALVQALGAAGVEAPLWLTNRNAVQ
ADGELVPAARLVGPRVVGIVSQPCWGLVDLVDAAASIRSLAAVLADPRGEEQV
AIRADGIPKARLVAPARATHEPLEAGTVLTGTGGIGLHARLWLSGAEHLVL
LGRGADAPGASSELBELTALGTGTIAACVADRARLEAVILAEAAEAGRTYSAVNH
AAGVSTPDLDTLEAFETIADYKRGVTNLDLDELCPDLDAFVLFSNAGVWVGSLA
SVAANAFLDGGTAPARASGAPVTSIANGWAGQNMAGDEGEGYLRSGOGLRAMPDRA
VEELHITLDGQTSVSVMDRRRFFELTAARHPLFDEIAGARARAROSEEGPALA
QRIALLCDGRERHLLHLLRAEVAVLHGDDAAIDRDFRFLDGLFDPKADDPPIAI
LAAVVGREATVVFDPHTITRLADHYLERLVGNAEQAQALVREVPKADDPPIAI
VGMARCGVGNPGLWELFIVGGDVAITEMPTDRGMDLALDPPDQORHGTYSIRHG
AFIDAGQDFAAPFGISPREALANDPQORVLETTWELFENAGIDTDSVRSQGLVFL
HAAVQGVQDAPVEDSEGLVLTGNSAVTSRVAIVYVLEGPATVDTACSSGLVFL
GSAQCSURDGDGCLAVAGGVSMAAGVETFEFSQGGVLAVDRCKASAEADGFLPE
GIVATVQRLSDGPAEGRGRLGVVVASAINQDQATNGLAAPSGVAQORVTRKAWAA
GTADVAVVVEHGTGRLGDPVEASALLATYKGRSSGSSGVLGSKVNSNIHQAAQA
SVAGVSTNAHVITAEPPPEPLPGPVGVIAAANSVYVLLSARTETALAAQRLLE
SAVDSVPLTALASALATGRALPRRALLAGDHLQRLGQLRVAEAGVAPGATGTA
SAGGVVVFVPGQAGQWEGMARGLLISVVPFAESIAECCDAVLSEVAGFSASEVLEQRPDA
```

PSLSEVDVQVLPFSVMVSLRWAGACGSPSAVIGHSGEIAAAVAVAGVLSLRDGV  
VVALAKALRAGKGVSLAAGERARALIAWEDRISVAAVNSPSSVVVSGDPEA  
LAELVACEDGVRKALIPVDYASHSRHVEIRITILADLIGISARRAAILPLYSTLHG  
ERDMGPRYWDNIQRVFEDEVAQSPDGHATFVMSPPHVLTAAVOEIAADVAI  
GSLHRDTAEELHIAELARAHVGVAVDMRNVPFAAPVVALPNYPFQRYWLAPEVSD  
QLADSRVDRWRLATTPVDLEGGFLVHSGAPESLTSAVEKAGGVVFPVPSADREALA  
ALREVPGEVAGVLSVHTGAAALHQSILGEAGVRAPLMLVTSRAVALGESEPVDEQ  
AMWGLGRVMELETPERWGLVDLPAPAPGDEGEAFVACLADGDEDOVAIRDHARYG  
RRLVRAPLGTRESSWEPAGTALVTGGTGALCGHVARHLAGCGVEDLVLVSRGVDA  
AAELEALVAGLPKTIITACDADRBLQSKLEELRGQRPVTVVHTAGVPSRPLH  
EIGLESVCAARLVCARLDLCPDAETVFLFSSGAGVMSANIGAYSAAAYLDALA  
HRRRAGRAATSVAMWAGMGMATDLEGLTRGLRPMAPDRAIRALHQLDNDGTC  
VSIADVWEAFVAGTAPRPLDELVTAVGAVPAVQAAPAREMTSQELLEFTSH  
VAAILGSGDPGTPAREASALDRGVQAGSVRSYLLDLAGLSDFEHGDSGDFS  
ADHIGQQSDGTAREASALDRGVQAGSVRSYLLDLAGLSDFEHGDSGDFS  
LDLVMDAPQGEVTVCCAGTAALSGPHEFTRLAGALRGIAVRAVPOPGYEGEPLP  
SSMAAVADQADAVIRTCQDRFVAVGAGALMAYALTELLORGHPRGVGLIDVY  
PPHQDAMNMLELTALFRETVMDDTRITAGAYDRLTGQWRPRETGLPTLLVS  
AGEPMGPWDDSWKPTFFEDTVAVPGDHTVAGDHTVQEHADATARIHIDAMLGGSN  
10867.12111  
/gene="eryA"

misc\_feature

Query Match 6.4%; Score 32; DB 1; Length 20444;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTCGACGGTGGAGCGCACGCCGCGG 413  
|||||  
Db 1033 GACGTCGACGGTGGAGCGCACGCCGCGG 1064  
|||||

RESULT 17  
AY283054 1209 bp DNA linear BCT 20-MAY-2003  
LOCUS  
DEFINITION Saccharopolyspora erythraea modular type 1 polyketide synthase  
gene, partial cds.  
ACCESSION AY283054  
VERSION AY283054.1 GI:30961866  
SOURCE  
ORGANISM Saccharopolyspora erythraea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.  
REFERENCE  
AUTHORS Nydegger, S.O., Schaeffer, P.M., Born, R., Stamford, N.P.J. and  
Haedener, A.  
TITLE Coding region of a ketoacyl synthase domain  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Nydegger, S.O., Schaeffer, P.M., Born, R., Stamford, N.P.J. and  
Haedener, A.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-2003) Department of Chemistry, University of  
Basel, St. Johanns-Ring 19, Basel CH-4056, Switzerland  
FEATURES  
source  
1. 1209  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1836"  
<1..>1209  
/note="ketoacyl synthase domain"  
/codon\_start=1  
/transl\_table=11  
/product="modular type 1 polyketide synthase"  
/protein\_id="AAP40326.1"  
/db\_xref="GI:30961867"

CDS  
1  
translation="VGMACRYPGGVTSPEDLWRLLAAGTDAVSAPPADRGWDRDLVP  
REGFLHDAEAFDSVSGISPREALANDPQORLLLEYSWEAFAGVDPLSVRSGETG  
VFAGVMYHDVATLMDVAFSELEGFTIGNNAASITARIATYFEGEAGVAVDTACSS  
LVSLHVAQALRAGCESMALGGVTVNATPGTFVEFARQGGASDGRCAKVFSDGADGT  
GWSBGAMLVLRLLRHGHPLVAVRVGSAVNDQGSNGLTAPNGPSQQRVLTQAL  
TNARLSTQDVVDVFAHGTGTALGDPIDRAQALLSAYGRDVSPLVIGSVKSNLHTQAA  
AGVAGVMKVLALRHGVVPTKLHVAEPSSHVDSGAVSLARDLVAMPEVDRPRAAV

SSFGMSGTNAH"

ORIGIN

Query Match 6.0%; Score 30; DB 1; Length 1209;  
Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGGCGAGTGTCCATGGCGTCCGCGGCGG 77  
|||||  
Db 519 CGGCGAGTGTCCATGGCGTCCGCGGCGG 548  
|||||

RESULT 18  
AF235504 77534 bp DNA linear BCT 19-JUL-2000  
LOCUS  
DEFINITION Streptomyces hygroscopicus var. ascomyceticus  
gene cluster, partial sequence.  
ACCESSION AF235504  
VERSION AF235504.1 GI:9280381  
KEYWORDS  
SOURCE  
ORGANISM Streptomyces hygroscopicus subsp. ascomyceticus  
Streptomyces hygroscopicus subsp. ascomyceticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
AUTHORS Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.  
TITLE The FK520 gene cluster of Streptomyces hygroscopicus var.  
ascomyceticus (ATCC 14891) contains genes for biosynthesis of  
unusual polyketide extender units  
JOURNAL Gene 251 (1), 81-90 (2000)  
MEDLINE 20323220  
PUBMED 10863099  
REFERENCE  
AUTHORS Reeves, C.D., Wu, K., Chung, L., Revill, P. and Katz, L.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2000) Kosan Biosciences Inc, 3832 Bay Center  
Place, Hayward, CA 94545, USA  
FEATURES  
source  
1. 77534  
/organism="Streptomyces hygroscopicus subsp.  
ascomyceticus"  
/mol\_type="genomic DNA"  
/variety="ascomyceticus"  
/db\_xref="taxon:132248"  
complement(412..1836)  
/gene="fkbW"  
complement(412..1836)  
/gene="fkbW"  
/codon\_start=1  
/transl\_table=11  
/product="fkbW"  
/protein\_id="AAF86380.1"  
/db\_xref="GI:9280382"  
/translation="MVPLUGRAPRRRTGLALTAANAVLLPSPATAPADPDG  
HCARLSQLRVAARQACDCLDELTAGTAATGHTDPADWAGLTPKNLATPTGVPGLQ  
IDGYFPTSTNTNMGHNDQFVIRLPDRNGLVAGTPGNGREQVANDRAIDWVL  
ARGVAAATDKGTGTFAYDGRSPGDVAEMNRLTQLTAAASVAAQRYHRLERT  
LVTMNSGVLVWQLENHPELYDGVDMWEGTLWRABGNLFTPLAALRDYPTAAG  
GPAAGQAREALLRAGFPADSEFLWPHVHEVYDITQRYREEDPGDGEKEACTPYC  
APGTVPGCADYDIATRSSGVHAAVERIALTGIGKPLITIQALSTCCCLPGGLRVR  
AVYRPPGRYRYRIEDGTHTDALVDTPDRLRPLPCHRSAPTALESMLTTGHRPP  
ANHTVVRPAGADQASLIAECLDG"  
complement(2020..3579)  
/gene="fkbU"  
complement(2020..3579)  
/gene="fkbU"  
/codon\_start=1  
/transl\_table=11  
/product="polyhydroxybutyrate depolymerase"  
/protein\_id="AAF86381.1"  
/db\_xref="GI:9280383"  
/translation="MQPFRGILTPFLPSSPPVCSLRPGRGVILTRLVAVLV  
LGAALLGAPATAAAGLAKEGLTKADLTVEADFGTNPGRINMYVYRPSLFAFPAVV



```
FALHGTQDAQGYADNSGLLSFADRYGFLLVFAETTSNNANRCFNWFQSSDNRGOG
EASTROMAATSVAGADPORVTITGLSAGGAMTSVMLATVPDVFQAGAVAGLPEG
CATDVSSAYLCNMGCTDLTAQWARRVDGYPSPWSGPMPRVAIHGDKDTIVARNAD
ELRQDQWAVGVSQTPDRTSVIGNSSTHREYSILAADGSVAVEVRKGLGHTGVPDGP
TGAQCGSTGAAYFLDSTCSSYWTQFFLGSASDPGLSPAPSLAATGATDTTISL
TKQFVDGATDVAVHRDGAQITTSATTSYTDGLRAGTSHTYVAARADADGKAGLSGA
VTAQTTGATACVCTAGNVAHYQAGRATTAGTYVYAKGSGQNGWLYNTEFTVTTLKESPT
GYFTVANDTCP"
complement (3969. .4496)
/gene="fkbk2"
complement (3969. .4496)
/gene="fkbk2"
/notes="transcription regulator"
/codon_start=1
/transl_table=11
/product="Fkbr2"
/protein_id="AAF86382.1"
/db_xref="GI:9280384"
/translation="MVEGQGRYPFEDFTTGQVYRHLGRTVLETDNSMLTLLTQNTAP
LHFDHYAGTTWGPVLVDSTFTLALVTGQSVTDISQHVMANLGMWRVRLNPNPVEGD
TIYSQSEVLNTRFSASRPDVGILSVRTIGYNQDGVIVITFERTLMVYRGGHGPFAAV
EPWDRSRRAAGLH"
complement (4595. .5488)
/gene="fkbk1"
complement (4595. .5488)
/gene="fkbk1"
/notes="transcription regulator"
/codon_start=1
/transl_table=11
/product="Fkbr1"
/protein_id="AAF86383.1"
/db_xref="GI:9280385"
/translation="WEIRLEWFTALAEHEVTTIABRLNISQPTLSRAIRLERRVG
VLFDRHQNLKINKEVFRRAHRAURASEVSGAEQRITALIDPDTGTVALGFLHSYG
TWLVPDLLAGRYRPAFTTFELRGDAADVDDNSRLDGLTSPRAGDDVVEWTPPL
RDEALCLIVPGHRLARRRVRTAELADEAFVALEPFGRLQITDRLCAAGFVPRV
LESTELSTRALVAAGLVAVAPLTGSPSPGDHAVAFLPDPQARRTVGVVTLAGGP
RAPVVARFVDVYRSRAHPV"
5601. .6818
/gene="fkbk"
5601. .6818
/gene="fkbk"
/notes="l-carnitine dehydratase"
/codon_start=1
/transl_table=11
/product="Fkbe"
/protein_id="AAF86384.1"
/db_xref="GI:9280386"
/translation="MSEPHRPRQERPGAPLSGLLVLSLEQAAVAFPAFTHLADLGAR
VIKIRPGSGDLARGYDRTVRGMSHFVWLNKESVOLDVRSPEGNRHLHALVDRAD
VLVQNLAQAGRLASATRSRGATEADHLRTYPATVPAATADKAYDVLVQCEAGL
VSIQTGTPPSKVLGSLADICAGMTAYSGLITALLKRAKTGRGQLEVMMLGEMW
GAEYTRYGGTAPARAGASHATIAYPGFTTDDGTQINLGNREMASFCGVLIQR
PGLCDDPRSGNADRVARHTELDVLSEVTGLTGBELVARLEASIAVARQTVREF
SEHPOLDRGRWAPFDSPVGALEGLIPPTVFHGEHPRGLRVPPELGEHTESVLAALAA
PHSADREEAGAAE"
6808. .8052
/gene="fkbk"
6808. .8052
/gene="fkbk"
/notes="similar to Rhizobium trifolii MatC; dicarboxylate
carrier"
/codon_start=1
/transl_table=11
/product="Fkbf"
/protein_id="AAF86385.1"
/db_xref="GI:9280387"
/translation="MPNELTGVLLIAAVFLVAGVGNLMLLALVATFLLGVVALDRT
PDEVLAGFPASMLVAVTFLPGIARVUNGTDVNLVRVAVRAGVAVPVPVLFGLA
ALLCATGAASPAVALVAPISVAVAFVHRIDIPLYAGLMAVNGAAGSAPSGLLGIV
HSALENHLPVSGGLFAGTAFNLAVAASVWLVRRLRLEPHDLDDTDTPEGDPAF
RPGAEHWMTLMAAULVGTVTLSDTGFALTALALLPRTSQATKETAMPVW
LLVCGIVTVYVALLQELGIVDSLCKMIAAIGTPELLAALVICVYGVVSAFASTTGILGA
```

```
IMLPSEFPLKSGAIGTTGMVMAALAAAAATVVDASPFSTNGALVWANAPERLPGVYQGL
LWMGAGYCALAPAAAAAFAFVA"
8156. .8824
/gene="fkbk"
8156. .8824
/gene="fkbk"
/notes="O-methyl transferase"
/codon_start=1
/transl_table=11
/product="Fkbg"
/protein_id="AAF86386.1"
/db_xref="GI:9280388"
/translation="MANQITLSDTLAYVRKVSRLRDEVLRLRAQTAELPGGVLVPV
QAERGQFLFLVRLTQARQVLEIGTYTGYSTLCARGLAGPGRVVTCDVMPKWEVGE
RYWEAGVADRIDVRIIGDARTVLTLGLDEAGAGPESFOMVFIDADKAGYPAYEALP
LVRRGGLIVDNTLFFGRVADEAVQDPPTVAVRELNALRDDRDRDLAAMLTADGVTL
LRKR"
complement (8802. .9884)
/gene="fkbh"
complement (8802. .9884)
/gene="fkbh"
/codon_start=1
/transl_table=11
/product="Fkbh"
/protein_id="AAF86387.1"
/db_xref="GI:9280389"
/translation="MTIVKLVLDLNTLWRGTVLEDDDEVVLTDEIREVITLDDRGI
LOAVASKNHDLAWERLERLGAEVFLARIGMGFKSQSVREIATELNFAPTTIAFID
DQPARAEVAFHLEVRVCYPAEQAATLISLPEFSPVSTVDSRRELVMYQAGFARDQA
REAYSGPDEDLRLDLSMTAPAGEEELSVEELTLRTSQNNATGVHYSADADLACF
TDPAHEVLVVTMGDRFGPHGAVGIILEKKPSTWHLKLLATSCRVVSGAGATILNWL
TDQARAGAHVADPRFRIDRNRMEIAYRFAGFADSDCPCVSEVAGASAAGVERLHLE
PSARPAPTTLTAAADIAPVTVSAG"
complement (9895. .10995)
/gene="fkbj"
complement (9895. .10995)
/gene="fkbj"
/notes="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="Fkbi"
/protein_id="AAF86388.1"
/db_xref="GI:9280390"
/translation="MPERDALITDLVGDRAAEWDTSGELPRDLLVRLQADGLLCAEVA
AEHGGILGSLRENGEFTAHVSLCSLSRVMTSQMAWTVORLGDAGORATFLKELT
SGKLAAGFSSERQAGSDLSAMETRVRLDGTAVDGHKVTWTTAAVADHLVVGLEQED
GSGAVVVPADITPGVVRVERVKPSCRAAGHADLHLDQVVRVAGVLAGSGASLPMLV
AASLAYGRKSVAGCVGIIIRACTAAVAHARTREOFGRLPGHQVLVAGHADIWTAEQ
IAARYCEYASDHWDEGSEPMVVPATILAKHVAERAAAGAAATAAQVLASAGAREGHVVE
RAYRDAKLMETIEGSSSEMCVNLQAHALAPA"
complement (10988. .11248)
/gene="fkbj"
complement (10988. .11248)
/gene="fkbj"
/notes="ACP"
/codon_start=1
/transl_table=11
/product="Fkbj"
/protein_id="AAF86389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAAHALVAFILKTKTRSDWPVDRDLFAEGGLTSLFAM
ELVVVLEKTFDVTQIPDLQLANFRFTVESMVALVHRLRAVDA"
complement (11245. .12093)
/gene="fkbk"
complement (11245. .12093)
/gene="fkbk"
/notes="acyl CoA dehydrogenase"
/codon_start=1
```

Query Match 5.8%; Score 29; DB 1; Length 77534;  
Best Local Similarity 100.0%; Pred. No. 0.0058;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

Qy 277 GCGGTCAACTCCGACGGCGGTCGAACGG 305
Db 37514 GCGGTCAACTCCGACGGCGGTCGAACGG 37488

RESULT 19
BD235937/c
LOCUS BD235937 77536 bp DNA linear PAT 17-JUL-2003
DEFINITION Polyketide synthase and recombinant DNA construction thereof.
ACCESSION BD235937
VERSION BD235937.1 GI:33045707
KEYWORDS JP 2002526114-A/1.
SOURCE Streptomyces hygroscopicus
ORGANISM Streptomyces hygroscopicus
REFERENCE 1. (bases 1 to 77536)
AUTHORS Reeves, C., Chu, D., Khosla, C., Santi, D. and Wu, X.
TITLE Polyketide synthase and recombinant DNA construction thereof
JOURNAL Patent: JP 2002526114-A 1 20-AUG-2002;
COMMENT KOSAN BIOSCIENCES INC
OS Streptomyces hygroscopicus
PN JP 2002526114-A/1
PD 20-AUG-2002
PF 01-OCT-1999 JP 2000574696
PR 02-OCT-1998 US 60/102748, 11-MAR-1999 US 60/123810 PR
PI CHRISTOPHER REEVES, DANIEL CHU, CHAITAN KHOSLA, DANIEL SANTI, KAI
PC C12N15/09, A61P25/00, A61P37/06, C07D498/18, C12N1/15, C12N1/19, PC
PC C12N5/10, C12N9/00, C12P19/62, A61K31/436, C12N15/00, C12N5/00 CC
Polyketide synthase and recombinant DNA construction thereof FH
Key Location/Qualifiers
FT CDS (52275) .. (71465) .

FEATURES
source
1. .77536
/organism="Streptomyces hygroscopicus"
/mol_type="genomic DNA"
/db_xref="taxon:1912"

ORIGIN
Query Match 5.8%; Score 29; DB 6; Length 77536;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GCGGTCAACTCCGACGGCGGTCGAACGG 305
Db 37516 GCGGTCAACTCCGACGGCGGTCGAACGG 37488

RESULT 20
AR271638/c
LOCUS AR271638 77536 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 1 from patent US 6503737.
ACCESSION AR271638
VERSION AR271638.1 GI:29703131
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1. (bases 1 to 77536)
AUTHORS Reeves, C., Chu, D., Khosla, C., Santi, D. and Wu, X.
TITLE Isolated nucleic acids relating to the fkbA gene within the FK-520
polyketide synthase gene cluster
JOURNAL Patent: US 6503737-A 1 07-JAN-2003;
FEATURES
source
1. .77536
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 5.8%; Score 29; DB 6; Length 77536;

```

```

Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GCGGTCAACTCCGACGGCGGTCGAACGG 305
Db 37516 GCGGTCAACTCCGACGGCGGTCGAACGG 37488

RESULT 21
AB070949
LOCUS AB070949 78210 bp DNA linear BCT 11-OCT-2001
DEFINITION Streptomyces avermitilis polyene macrolide biosynthetic gene
cluster.
ACCESSION AB070949
VERSION AB070949.1 GI:15824136
KEYWORDS Streptomyces avermitilis
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
REFERENCE 1. (bases 1 to 78210)
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
PUBMED 11572948
REFERENCE 2. (bases 1 to 78210)
AUTHORS Ikeda, H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)

FEATURES
source
1. .78210
Location/Qualifiers
/organism="Streptomyces avermitilis"
/mol_type="genomic DNA"
/db_xref="taxon:33903"
/note="Actinobacteridae"
complement(1..1254)
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="BAB69301.1"
/db_xref="GI:15824137"
/translation="MGEALITLCSTRSLAATGVLGGVLAATLGGTSMNATSGNCA
GNHPTSOARINQAGAEADARIEITPGQTSVGTNHPVSVTVTSKLTMTVA
TGAEIPGTLSTGTSWKPNGRLERATKYQIAAAGADVKGRSTTGTATTVSPANDFI
CHLSPDGSTVGVMPTVNFDAISDKAAQVSEIRVSSSGQGVGVGHFNDHRVDFR
PENYRPGSTVTVTLNRHGQKTVFKIGRSQISTVDKATKQMTVVRDGTIKTIPIS
SGSPHEFTYNGQWVISEKFRQIHNWASVGLTEKNGKPSVDIKAPVHMLTDSGTFI
HGNYWGTDSYFGEANTSHGCVGLKQVGRDGHGKQPAAFDFDHSIVGDIIVKNSSEKT
MLAPDNLGSLDNWMSSEWVAGSPAT"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAB69302.1"
/db_xref="GI:15824138"
/translation="MPVEFTSHDQVSRYSRYPASAPTPELGFQFFFLDAKALEASRSKR
APANRLGAWQWGVYMLGVEPTEDLSVPEVVVREAEQOLGQVDPGFPAAYCARQNR
YEHAWERDAYRESPAARQVREFLATRWASWEGPRALFDRAGVGLVNRVLLPG
ITTLTRLVAEYAEIRMEAVAMFTAGPSTDIKELRVSVQSVQRWRSGLAGDRPGRG
GVSGSADRLAFSEIRMEAVAMFTAGPSTDIKELRVSVQSVQRWRSGLAGDRPGRG
AVPOPVSVELSETLPAALDERAKGAFPSWTIRYLHHDFFPIFGQRYAYKWSQGLHSF
GRSAQGVDDQ"
.2935.. .21081
/gene="pteA1"

```

2935..*21081	VTGGDLAQSAVCLVRSQTEPNDRILVIVETSSDAHAGAGSGVDDADLADLAVLPGLIA	Query Match	5.8%;	Score 29;	DB 1;	Length 78210;
/gene="pteA2"	SSEPOAHADDEVRVRLVRAVVDPAEPELPGSGGTGVLITGASGGGLGLPARHLVA	Best local Similarity	100.0%;	Pred. No. 0.0057;		
/codon_start=1	EHGVRSLIVSRGGDAPGAELTADITAGADVTAAACDVADAAVRAALLAGGQSL	Matches 29;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
/transl_table=11	SAIITHAGVLDDGIIGSLTPELDAVFRPKVDAAALNDLAGELSGDLSAFLPSSVA					
/product="modular polyketide synthase"	RTIGTGPQANYARANTFLDALAEHRAAGORATSLAWGLWQDSSEAMTGLDHTLT					
/protein_id="BAB69303.1"	GI10824139					
/db_xref="GI:15824139"	NLNTVPESALGAKEGGLSPDCRFTEDARANGYVRGGGIVLKLPLARAVADGDPV					
/translation="MVSELSNATGNRDCSASDAGTAVIGLACRUPGAADPAAFWR	CVTRAGLVNDGGDGLTVPQSGOEQLVLAVERAGNPDVYVELHGTGTGVGD					
	PIEAAAGLVGTGEPGRPLRVGSATKNVGLHGAAGITGLLKAVLSINRELPSAL					
	NFGTPNPAIPLDLNRPLRVQTESHWAEDGRLLAGVSSFCMGGTNCHIVLASGAAT					
	AGEGATAGGAAFTWPLSARTAGSLRAQAALAHTEAPLGLALPDVMSLACRTA					
	FEHRAVVEVRDFTFLRELSTGGIDDAALTGRAGPRGELAFISQCGQRAGMR					
	ELASVPAPASALDEVCAHLDPLLPRLPREVMEFAPVGTGEAAEDQTLTQTSFAVE					
	VALPRLSMDITPDVLMHGSITGELAAAHVGLSLADACALVAARGLMQALPAGGA					
	MTAVQAAEDEVARIGERTDRVSIAALNGPDSVVVSGDEDLATEADAFALGRKTSR					
	LRVSHATSPHMEPLAFRVAEGLTEHAPRIPIVNSVTGRLSEEPSGEGCTAAVY					
	LVHVEARVFDAGVARLDEQGVRTYIELPGCGVILASMARAGANDSLSLVFPAIRTRPE					
	AQALAAVASHVHVEPDMDALFGGGRGARKKVALPIAFERQVLDGDDTTSVPAT					
	TGWAAPQPAIATSDPAPGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR					
	QVETEWTKOLFDSLSVELNRQNSEATGLPLSGLLFDHPTPAALAHHLAAQSLGG					
	GDAAEAATVADPHEPIALVAMSKLPGVSSPEDLWLLASGGDGI SGFTDRGW					
	VELTYDPEGTGPKSTHGGFLHEAAREFDAGFGISPREAAIDPQORLVLETAWEA					
	FERAGIDPATLGRSRAGVFIGATQDYGRPLHEPADGVGGLLTGTTISVAGRVAYS					
	FGLEGPAVTDTCSSSIVALLHVAQUSRSGETCWALAGGVTVMTGPMVFBSQRG					
	LSADCRKAFAGADGTGWAEGVGLLVERLSDARRNGHOVLAVVRGSVNDQDASNG					
	LTPANGPQORVVRQALAAAGLAPADVAEAGTGTGLGDIIEAEALITAYGQGRDG					
	LDTEAVMLKSNIGHTPAAVSSSGISCTNAHVILKMMVMRHGVLPRTLHVDPTEPHVMSASGVR					
	KSQVALREQARLLSFLDABGVGAGVRPDEVALSATSEARVLERAAVVGGDLAELR					
	QGLEALASGAAPVGGVAGVRGFLFGSGQSQRISQRMGRELIAAYVPVPAAYDEYCA					
	FLDAPVDSSENLGTQTPALFAVEFVALFRLISLWSGIRPDYVAGHSVGEIAAAHVA					
	GVLSDEDAKLVASRAALMQALPDGAMVAQVATEDVHLHLDGTVGIAAVNGPSRV					
	VSGEADATVAGEIFAGRGKTSRLKVSHPAFSLMDPMLPEFAAVVRGLTSESRIP					
	VVSNLTGLAEVTVPEYVHVREARVADGIIETLGLGVTVFVEPGCGVLSALAQQ					
	CFDDGIDIVTVVPLRADRPPEVLAIVAGRLHTHGISPDNQVLPFGARVDLPVTAFO					
	RRERYMDAPATAGADAMAEFWSDESORASIGALLGLEPTELDVVASKLAMR					
	RRERQSTTDDRYRITWQFDLDDMVRAPSGTWLVYVSEETAWTEAIRAGLTELGVAL					
	VFPAGEQTDRLARAFADAGOEADGILFAAAQGEAGANALQRLVILVQALGDAGI					
	EAPLWCLTSAGVSTGADLITDVPVSAQLMGLRIVALEQOPQRMGVLDPVPAELDSRAL					
	ERLAGLIAQADEDLAVRAQSVYVGRPLVRAPLSAAPADAPWSRGTVVACDADRDALEL					
	QVARILAGAEHLVITSRGPDPGAAELKAELELGAQTVTVACDADRDALEL					
	GRYINAVVHTAGVLDGGLVESITPERLNRHVLRPKVDAAHLHLELTRDQDLDAFVIF					
	SSMTGVNGGQGYGAANAFDLALAEHRAQGPALAVANGSWADGGMADGADGHL					
	RRRGVRAIALPAISVLHGALTHEGTSMTVDVDFERFVAFVGTGRCPLLGVPEAR					
	EALAEARAQSPVFPVSALVORLLGATPGEQRIILDLIVREQAATVLTGHTKGAFEA					
	DRTPRETGDSLTAVELRHLNLTALGLRPTTLVFDHPTPTALARHLGELLGRHGLE					
	TAAQAATAAADAEPIALVQMSCLPGGVSPEDLWELISSGRDAMSGPTDRGWID					
	ALYDPPDPRGKTSREGGFIQADRFDAGLFGISPREALAMPQORLLLETAWEAFE					
	RAGIALVASRASVTGFIITNGQDYANGLENAPETEIGYALTAKAASVSVSGSYTFG					
	LEGPAVTATCSSILVALHAAQALSGSCSMALVGGVTVMTPTDLFVFSRQRLS					
	PDGRCKAFAGADGTGWEVGVLVRLSDARRNGHOVLAVVRGSVNDQDASNGIT					
	APNGPSQORVIRQALASAGITPADVDAVEAHGTGRLGDIIEAQALILATYGQSHSDR					
	TEVQVPGTTPRRRAVSSFGIGTNAHTIIIEAPAAAAAPAEHRRPVPVFWLSAK					
	SEVLAQARALLAFATDDYSPVDAGFSATTSALEHRAVITGSDPDAELRAGLEALA					
	AGPPAANVVAHRAHSAQKVGFLFGSGQSQRISQRMGRELIAAYVPVPAAYDEVCAILLAT					
	VYDVAETHGTGTOPALFAVEFVALFRLISWGRIPDYVAGHSVGEIAAAHVAHVLISL					
	EDAKLVASRAALMKALPAGGVMAVQAEDMLPHLIDRAGITAAINGPSPVTSN					
	DAVTAIAEVFAAQGKTSRLKVSHPAFSLMDPMLPEFAEVVRLTISPEWIPVSNL					
	TGRLTESYTPYVHVREARVADGVITLHELGVTVFVEIPGGVLSUALVQGLDDG					
	LDITVITIPALRADPEQALATAYLAQHVSGAEVNDWHAFPPCARREVELPTVAFQREYV					
	LDAPASAGDMRAVQGAHPHLLAQAAPLADSVGHLLTGLRSLKHTHPWLVHVAHSGVA					
	LIPCTAFAVELAITAADAVGCDLLELTLETPLLVPERGGVALQVHVPVAGSDGSRPRT					
	VHSQIDGDGDAYAHEDDSERAWHASGVLTATGTPVDGSI,TATPPVAGAPBVDGTY					
	DRUAGMALDYGVPFGQRLAARWSGDDFFAEVELPGBERTDAGATGLHPALFDAAHTV					
	WLGAVEPEAGTGNPGLPFAWSGVRLAAAGAKWVKVSRAGTSTVSLVLDGTGPF					
	VAEVDSITLTPVADQLPFAWSGSDDAVFLGKMTFVSLPSPADGNRIETSLDIALRAT					
	DDPADTEADAAAAADA VVVPCTGTGADLATRVREVTHAVILELVQVWMLABERFGRVLV					

Qy 382 GACGTGACCGGTGGAGCGCACGCCAC 410  
 Db 47502 GACGTGACCGGTGGAGCGCACGCCAC 47530

```
RESULT 22
AF357202      113193 bp      DNA      linear      BCT 04-NOV-2002
LOCUS      Streptomyces nodosus amphotericin biosynthetic gene cluster,
DEFINITION      complete sequence.
ACCESSION      AF357202
VERSION      AF357202.1  GI:14794889
KEYWORDS
SOURCE      Streptomyces nodosus
ORGANISM      Streptomyces nodosus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomycetes.
REFERENCE      1 (bases 1 to 113193)
AUTHORS      Caffrey, P., Lynch, S., Flood, E., Finnan, S. and O'Liinn, M.
TITLE      Amphotericin biosynthesis in Streptomyces nodosus: deductions from
            analysis of polyketide synthase and late genes
JOURNAL      Chem. Biol. 8 (7), 713-723 (2001)
MEDLINE      21344785
PubMed      11451671
REFERENCE      2 (bases 1 to 113193)
AUTHORS      Caffrey, P., Lynch, S., Flood, E., Finnan, S. M. and O'Liinn, M.
TITLE      Direct Submission
JOURNAL      Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             source
1..113193
/organism="Streptomyces nodosus"
/mol_type="genomic DNA"
/db_xref="taxon:40318"
complement(4..1824)
/gene="amphG"
complement(4..1824)
/gene="amphH"
note="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="AmphG"
/protein_id="AAK73498.1"
/db_xref="GI:14794890"
/translation="MAPDVEEHEEERESEQPVRLAALLRPHRRSVGLALTAVGVGI
LNAFPLRLGLLIDGLVCHGPGVDFGALGRLLMILLVLYVASVFLVQN
LVASVVRLLIDLRHAEOKLARLPLRVDFRKPACETISRGTDVNDLQOITL
DLISSVFLVIMLSLIIISPLAGVLLSVPSGLLAWSKRAPOQVAAOWMSGK
LTAHEMCAGHALVAFDRAREQRFBERRNEAVTAGSGAQFASGLPEVPMFYAN
LGVAVAVGVKAVNGSVLGLDQVAFILYARFSQPIVEIASVAGRLQSGVASARV
FTLLDAPEQEPDFPLAVRVEGRVEFQVSPRSPDPLIBGLSLVPEPGSTAVV
GPSGAGKTTVALLMEFYEDISGRILLDGTDAAMNDDLRSLRGLVQDTWLFKGTI
AENIAYGSPGATRADIVAAATYADRFTLTSQGYDTLDDSGGVSAGEKOLITVA
RAPLAPAVILDEATSSVDTRTELLIQRAMNLRPGRTSFVIAHRLSTIRDAVIV
MKSGRIVEQTHDQLDAQAYARLHARADAADAVTVG"
complement(1805..3628)
/gene="amphH"
complement(1805..3628)
/gene="amphH"
note="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="AmphH"
/protein_id="AAK73499.1"
/db_xref="GI:14794891"
/translation="MARSRWPTFPQAAAGRPQCHSVLLRLMRQLRYPAGSVVA
LVHLVQLLGLTLLGLAALIDEGVRRHSDRIGTIGATMAVALVQIAALGAA
LGARTSALGRDLRSARVRLDPSAREIRFGTSPSLTSSVNDVQVQNLQAGUGI
FVAAPLCLGSLALQDVTALILVPMVLVAVCFGLLSRMAALVQLQOITLRI
GRILRRIIVRVRSRQAICAGEFRTEELLSGLVGRLLIAMLPSVLLMNL
FTLGLLVGARRISQWQIGALSFLSYLSLILMSVYMLAFVFNVPARVCAERTI
EVLOQETDVVPASPRMAGPAGVGVHAFRFPYGBNAVRLSLTLRPGERVAVI
GSTSGSKTLLHLILVLDVAVAGEVIRGGTIDVRELDPSVMAAAVGYVQFPLFAGV
ASNNRFRPATDELMEVIRIAQHEGFVRLGLDTEIAQGGTVTSGGQRQRIAR
ALLRPAIYLPDSSFSALDQSTEAARKALVPYEGATVITVAQRVASVRDADRIVL
DOGGIATGTHDALLRDSPTYREIATLSQRTREETAHGAGRS"
3840..4874
gene
```

```
CDS
/gene="amphDIII"
3840..4874
/gene="amphDIII"
note="mycosamine biosynthesis protein"
/codon_start=1
/transl_table=11
/product="AmphDIII"
/protein_id="AAK73500.1"
/db_xref="GI:14794892"
/translation="MPKRALITGITGQDSYLAELHLLSQGYQVWGLIRGQAMPRKFRV
SRLASLSFVGDMLDQSLVAVDQVDPVYNLGAISFVPMVQQAELVTVNGMG
VIRLEAIRVWSGLSMSTAGTQIRFYQASSEMFCVAKETPQRETVLHPHPRPYG
AAKAGHFITRNYRESFGYAVSGLMNFHESFRRCQEFVTRKISLAVARIKLGLODKL
ALGNMDAVRDMGYAGYAVSGLMNFHESFRRCQEFVTRKISLAVARIKLGLODKL
KDYVVDQILVRAEVEVLCDSSKAQQLGKPSVDFQELMRMVDLADLASVSRQNE
LDDILLASHW"
5042..33574
/gene="amphI"
5042..33574
/gene="amphI"
note="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/transl_table=11
/product="AmphI"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="MDNQKLRDYLKATADLRARRRRVGELESQSEPIAVGMTCR
YPGVSSPEDLRWYVAGEVNGVPTPTDRGMLEALASSPTSRGSGFLHDPADDF
AGSPREAVMDPQQRVILSEAWERAGINTSVKSGRTGVFIGMAQDYVGPADG
AEGFQLTGNTGSLVSYTFGTVPVDTACSSLVAVHLATQALRAGECTIAL
AGGVTVMSGPTFIEMGRQGLSVDRGCRSFGDTADGTGAEVGLVLERLSDAIRN
GREILAVRGTVANODGAGNGLTAPNGPQQVIAEQALVARLSAGDIDVEAHGTGT
TIGDPVEAQALLATYQQRDEKPLILGSKNISHTOAAAGVAGVKNWAMRHGVL
PRILADETRHVDASQGVRLVTEWTPACAPRAAVSSFGISGTAHTVIEQAP
EPEADPEDDAPSTPAAVTGVLVLLSGRSPVLRQAALLTLTGTPPPADLAV
SIATTRTAFREHRLVLIADLPETLGRITAIAGTDPAVLDTVGTARTELRLVLT
GOGAORLGAAGRELAAFPVFAALDAALDAFTPLDVLKRVKVLGADRLDREVAQ
PALFAVEVALYRLLESFVKPHLAGHSVGEIAAAHVAGVSLDDAATVAAARGLMQ
ALPEGMAVAVQASDEVAFLAGHEDVLAANVGPASVLSGDDTITVELAARLAA
DKGRTSLVSHAFSPLMAPLMDFRNVVEGLTHSLPLVSDVDTGRTAVALTSA
PIYWDHVEQVAFADGIDWLKHADVTAFLELGDSVLSNMAQCLDAGSDALTVA
LREGEDEFTTAAALHTQGTALHWDACFTGTGARTDLPITTYTQRRWPRVAVQ
GAADLRSGVGLAAHPLLSAAVSLADSEGALTGRISSLHFWLADHTVGRALLPT
AFLELAVRAGDEVGCDRDELTLAAPLVPEQGVQVQWLMGNPDASGRSVTVGRP
DADEAPWTSHTAGVLSASRTSDFDATVMPADAETLPVDGLYERLAGGFGYGLF
QGLRAWRGDEFAVFLPESGHTDAESGHLPHALLDSALHAASFVLDLRAAGGLP
FSWEGVLSHAGSATTLVRILAPAGDAVAIADDSGQLVLSADSLIRAVAREIDA
AAALVRDALFRLDWVPTVAASGTAALVGEDPFGRLALPQSGDLAHFIDLALAA
DGAVPDTVLLPTGTGPDADPVTAAHRAATEALAAVRLTQDERFAASRLAVTRGA
TTGHDPAAAVWGLVRSQASNGRFLVLDADQDTPALPAAALTSEEPOLAVRGE
LRAALVRPASTAEAVPAFGGEGAVLTGCTGLGALVLAHLVABHGVRELVLVSR
GGAAGAAELVAELAESGARATVACDTRAAVLAELVAHPVSAVHVSAGVLDGNGV
GLTLPRLTTLVRPKVDAWNLHEATRDLDLKAFVLPSVAVGLSGQAGVYAGNAF
LDLAAHRRRAAGPLGSLGWPMWEGTGGTGGSEDDLRMRARAGTALIVEQGLALL
DAALDGDAAALFVRLDLSVIRAQGVPLRLSLIRSRRAAVAGATAGGLAQRUA
RLDASGRDELVLVLDQVALVGHATGAIEDAGRAPRELGFDSLTALEVLNRLNTV
GLLPATLVDPDYPTVSHLSYVLDLLELGEVAVVQRTAAVADDPTIVTGMACRYP
CGVTSPEDLWRLVTEGTDVSGFPVNRGVDENLYHPDPDHPGTAYTRSGGLHEAGE
FDPFGFGMSPREALATDSQORLLLEASWEALERAGIDPVGRLSGATGVFAGVYSDYS
AMLSPEFGFGGSGSSPLASGRVSYTLGLEPFAVTVDTACSSLVAMHMAQALS
GETSLAAGGVTVMSPTPAVDFPARQGLSPDGRCKAFSDSADGVSGVGMVLVLR
QSDAIRNGHOILAVVRGSAVNODASNGILTAPNGSQQRVIRQALASGGLSAGVDVY
BAHGTGTTGLDPIEAQALLATYGRDRPROPLILGSKNISHTOAAAGVAGVKNW
TVELQAPGTVPAAPAPERTAGVAPLLISGRTDRALRQAPVAVEGAARAGVAMFSG
SMRGVHLPRTLHVDASSHVDVTEGAVELLTEQATPETCRPRRAVSSFGISSTNVH
GHSLATRSRFFERRAVIAQDREGLASGLAAGDPAARVEGAARAGVAMFSG
QSQORAMGRELVEYTPFAAFADVEVCAALDPLDLRLREVFAAEGSEEAALDRTG
WTOPALFVAVLYRVESWGRADFVTHSICIEIAAAHAGVTLLQDAARLVAARAT
LMALPSGGANVAVQATEEVAPLIGHLSVAAVNGPTSVVSGDEGPVAAAEFSG
RGRRTKRLRVSHAFSPHMDAMLDLAFRTVAETISYAPRPLVSDLTGRADDAEFTV
VLRVRRVRAVAFADCVTRLRDAGATFLLELGDPLGLTMAEDTLGDERYDNTALV
PLLRADRPESAAATAARLIQIHGVDLDWTAYLAGTGARRVDLPTFAQHAHYWQPL
```

SAAPSAGDPAOKUWAVERGDAEELAAVLGDESDLTPLDLSLLPALSSWRRGNOEK  
ALDUTKRYVWTRLSKPAAPVLDGTWLLVSSDADDETELDGLAEALGARVVR  
RLVLDQADRAVLARLDATENADNTAQVLSVLPDDEFTDGPAGTQOGLALITAIIV  
QALADTCAGHRLWATRGAVSTGPADPVTHTPAQATWGMGRGVALHEPRWGLGVDLP  
ADPDGAGORLWATRGAVSTGPADPVTHTPAQATWGMGRGVALHEPRWGLGVDLP  
VLTGTGGTGAAGLWATRGAVSTGPADPVTHTPAQATWGMGRGVALHEPRWGLGVDLP  
DVADRALAVLGLSDADPLTGVVHTAGVGHYGLDPLTAPAEFAGLTAALAGAHL  
DNLGDELDFFLFGSIAGVWGSQOSAYGAANAVLDALARRAGLAAITAMGP  
WGTGWAADDAVSTGRLGLGLDPPALTEMRRAVRQDVTVTVDVADWTZYAPLF  
TSAPGALISDLPEVRAALAEENTPADTDAEISEIVQRVSEPEQLKLLDILVTEKA  
TVLHGSAGAVPEPRAFRETGFDSITAVELRKHGATGLSLPSTWVTVLSTPLAQL  
VLEAWMGVSVLAVGVPATGFTDDEPIALIGMSCRYGPGVSSPEQLWDLVLSGDAIT  
DFVNRGWTAGLVDPDPDPHTTSTQGGFLHAEDEFDMFFGFSPREALVMDPOOR  
LLLETWEAFERAGLTPDTRLSLTGTGFISSVQVCEMGAGDGAELHVTGTSPLVS  
GRLAYVGLGRLVTPACSSLLVALHQAQLRNGESNLAVAGATVMTTNAFVA  
FSRQALAQDRCRCAFESDGMTLAEGVGLIVLVERLSDARRNGHPVLAVIRGSAIQ  
DGSNGLSAPNGVQSVIRIQALANARVAFCEIDLEAHGTGTPPLVIRQAQALFAY  
GRTRTPETALLGSGVKNIGHSOSAGVASTIKMVALRHGVMPQTHADEPESHVDW  
SPGTVELLGTWMPOTGRPRRAVSVSGISGTHAVHILQEETAPAEDEQLAPPL  
FVAGVVPILLSARGAALREQADRLTHLVTADPAAREPIDIGLSLATSRLPHEHRAV  
VVPAGTDPLEALRAVADGPGVARGVADVAGRTVFPFGQSQWAGMCAQLLDES  
PVEABRTAECAALAEFTDNLIDVLRGAGAPTLERVVVQPSFPAWGLAAVRA  
QGVPEADVHSGEIAAAVYVSGLSLRDCAVVTLRQAGIRSLAGRGWMSVALPV  
AEVEARLEAFEGRSVVAEAGPRSSVAGPEALDELHAQITABEIRARVAVDYASH  
SPVEDIDHELELLEAEVAPRTSIPFSTVTGDMVTDVMDGYWVRSRGLRFLLSAA  
AVRDLIAADHRAFTLEVSHSVLANVQDMIDDAAGVAGVAGCTLRDNGGLDRFLLSAA  
EVFVGVQVDAWAFEGTGASRVLDPTFAFOHENLWMAAFAEATVTAADPEDAFWTA  
VEDGDSALTAALGTDBDSVAALPALSWSRRKERSIVDSWRYRTPKRWTKLPQR  
TLDTGTLVLDGVDVDDTVDALEALTEGAEVRLVLDSCDRAVIRLTPKRWTKLPQR  
VSVLGAETGATGVLGVALTVLQALDAGIDTLPWALTRGAVSTGSDKVT  
APVQAVTGTGATWALECPGWSGVVDPETLDAAGORLAALVLAGLGGDQDQALRS  
SGVTRIVRADAPDSARDKPGTTLTGGSGTAPHLARLWAGVAGHVLVSR  
RPEAPGAELRLEAEERGETTLAQCIDTRDAVALLESLSKEGRTVTVVHTAT  
TELEHTLDDDFDLVAAKTAQILDELLDEELDFVLSSTAGWGSAGHAAV  
VAGNAYLAALAEHRRAGLTLASLWSGVALDQLGRVDPQMRSLGFPMDPOLAS  
GLKALDDDEOIVAVADVMTYHPVTTSARPTPLFDEVEVQRLTAAEGSAGPAR  
GEFAALLAPAEHQHKLLETRTEASVGLSSAEDLTQDRAVDVFGSLGTAVGL  
RNRLASVTGLTLPSTWPDYENPALAGFLHSELADVHAGAVATAGAPVDDPPIAL  
VMSCRYPGGTSIAQLMRVSLSEVDVAVFPADGWDAREALYDPPDASGRVTSVQ  
GFLRDVADFPFGFISPREALSDMPQRLLETAWEFENAGLDPVQAQRSGTGTFI  
GASQDYGAAYVGGSGSGHMTGSLPSVLSRVSYLFLGEPVATLTCSSLVAI  
HLACQSLRGESTLALAGASIMSTPMSFIGFSRORALAEGRCKAYAGAGDGMTLAE  
GVGLILRELSDARENEVLAVINGSVNOGASNLTPAGNSQOQVIRQALANAG  
VEANDIVLEGHTGTALGDLPEAQALPATYKORDPERPVLGSKVKNIGHTOMASG  
VASIILKHALREGVAPKSLHDQSTHVDWSSGTIQLLTERTEPTEGRPRRAVVS  
FGLSGTHTVLEQAPADAFAEDTPAPRDALVPLVSGRGEALRAQAQALLDLA  
ERPGHTPDLAFSLATSRAALEHRAVAVADDHEALVRLGLTARDLPGAGLVQGRTR  
GRTPFLTGQSORLGMRELYERHPVFPADALDAVLRIDGTTERPLRDLVFAEGSQ  
DAALLHRTGYAQPALFALEVALFRLLESNGVTPDYLAGHSVGEIAAAHVAGVLDLDA  
CTLVAAHRLQALPEGGMVMALEAEDEVLPHLEGLADQVSAVAVNGPRSVVAGEE  
EPVLAAAHFAEQRKTRKLRVSHAFHPLMDPMLDDPAVARALTYHAPSVPVSNV  
TGTTLAEPQVCTTADVSHVRSVAFADGWLSTQGGVQTPLELGPVGLSGMARES  
LTDASRTALLPTLRDREPEQALVAVAAHAGFDVDWTFWFGSGARRVALPYAF  
QSRVWYPTTAAIGTAPAPAGDAEFWAHVAEDVAVSLTASGLDDATVTVAMVALLT  
AWORRGOSALDSNRYVTKPRGAGCAAPTGRVLPVPAEHREDEATAAADAEEA  
ALATATVREVVTGBRALAARLTERADGDTFGVLSLALAPGAGHGAAPALTTLT  
ATLQALQAGIARIDALWNLITRGAVAGVRSQVTAPEQAARVWGLFRAALLELPARYGGS

Query Match 5.8%; Score 29; DB 1; Length 113193;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGACCGGCTGTCGAGGGCGTCGG 203  
|||||  
DB 51354 GACGGACCGGCTGTCGAGGGCGTCGG 51382

RESULT 23  
AX703543  
LOCUS 113193 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 1 from Patent WO02097082.  
ACCESSION AX703543  
VERSION AX703543.1 GI:29538472

KEYWORDS  
SOURCE

Streptomyces nodosus  
Streptomyces nodosus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE

1 Caffrey, J.P.  
Engineered biosynthesis of novel polyenes  
Patent: WO 02097082-A 1 05-DEC-2002;  
UNIVERSITY COLLEGE DUBLIN (IE)

FEATURES

source  
1. 113193  
/organism="Streptomyces nodosus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:40318"

ORIGIN

Query Match 5.8%; Score 29; DB 6; Length 113193;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGACCGGCTGTCGAGGGCGTCGG 203  
|||||  
DB 51354 GACGGACCGGCTGTCGAGGGCGTCGG 51382

RESULT 24

AX703543/c 138203 bp DNA linear BCT 06-DEC-2003  
LOCUS Streptomyces sp. FR-008 heptane macrolide complex synthesis gene  
DEFINITION cluster, complete sequence.  
ACCESSION AV310323  
VERSION AV310323.1 GI:34766435

KEYWORDS

Streptomyces sp. FR-008  
Streptomyces sp. FR-008  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE

1 Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and  
Deng, Z.  
Organizational and Mutational Analysis of a Complete  
FR-008/candididin Gene Cluster Encoding a Structurally Related  
Polyene Complex  
Chem. Biol. 10 (11), 1065-1076 (2003)

JOURNAL

14652074  
2 (bases 1 to 138203)  
Chen, S., Huang, X., Zhou, X., He, J., Jeong, K.J., Lee, S.Y. and  
Deng, Z.

TITLE

Direct Submission  
Submitted (30-MAY-2003) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

FEATURES

source  
1. 138203  
/organism="Streptomyces sp. FR-008"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:206662"  
complement(574..1950)  
/gene="fscQ"  
complement(574..1950)  
/gene="fscQ"  
/note="putative FAD-dependent monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/product="fscQ"  
/protein\_id="AAQ82549.1"  
/db\_xref="GI:34766436"  
/translation="WPSTCAPTSTGFRGCHRRKRTGRGRIAVPTPCGKVRKGH  
QAKKRVYMRGSGVAVVGGSTAGCAALAAASRGGAERTVFERADARLDRGVGIGV  
HDAVRLADAGYMADEMPALNRRVWSDGDAHGHGAIQTGPPFPAYNNGSLWS  
ELRRVPEATYRAGAKAEVQDADGVTVRLADGPEPFDIVIGADGYSVUREAMY  
PCTGAAYAGYIGWRGTSPTDVSGLPSDGLDAHNTTFPGHGMAYRIFDGGGHRNLWL

gene

CDS



/protein\_id="AA082556.1"  
 /db\_xref="GI:34766443"  
 /translation="MATHFVSPALDGRLEVYSDVSGNWISSQGVYRREERFA  
 EWNGVHGVCSSGTAITLALRALNIGPGEVIVPEFTVWASAWATYTGAPVFD  
 CGDLNIDVTIRIEKIPTFRVMPVHYGRCDMDVMDLQYINURVDSABAGV  
 VRPGDIACFIITRAGEGGVCLTDPRLAQLAHRAWATDHGFLHKKLAY  
 NYRTMOCGVALAQTREILATREIEARIDAGLKLPGITLMPARDVLYMYDLR  
 AEEREERHLHDARGIETLFFKPMRSQPGYLDVPWFTLNAHRSDEGLYPLTHGLT  
 AADQYITGAVRDPFRAG"

Query Match 5.8%; Score 29; DB 1; Length 138203;  
 Best Local Similarity 100.0%; Pred. No. 0.0054;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 CGACGGCCGCTGCAGGCGTTCGCCGACG 169  
 Db 116232 CGACGGCCGCTGCAGGCGTTCGCCGACG 116204

RESULT 25  
 AP005022/c  
 LOCUS 300425 bp DNA linear BCT 10-MAY-2003  
 DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 2/30.

ACCESSION AP005022 BA000030  
 VERSION AP005022.1 GI:29603894

KEYWORDS Streptomyces avermitilis MA-4680

SOURCE Streptomyces avermitilis MA-4680  
 ORGANISM Bacteria; Actinobacteria; Actinomycetales; Streptomyces; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
 AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
 TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
 MEDLINE 21477403  
 PUBMED 11572948

REFERENCE 2  
 AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S.  
 TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis  
 JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)  
 MEDLINE 22608306  
 PUBMED 12692562

REFERENCE 3  
 AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T., Kishida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y., and Hattori, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
 COMMENT (E-mail:bioente.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)

COMMENT This work was done in collaboration with Haruo Ikeda (\*1), Jun Ishikawa (\*2), Akiharu Hanamoto (\*3), Chigusa Takahashi (\*3), Mayumi Shinose (\*3), Hiroshi Horikawa (\*4), Hidekazu Nakazawa (\*4), Tomomi Osone (\*4), Norihiro Kushida (\*4), Hisashi Kikuchi (\*4), Tadayoshi Shiba (\*5), Yoshiyuki Sakaki (\*6, \*7), Masahira Hattori (\*1, \*7) and Satoshi Omura (\*1, \*3).  
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa

\*1 Kitasato Institute for Life Sciences, Kitasato University  
 \*2 National Institute of Infectious Diseases  
 \*3 The Kitasato Institute  
 \*4 National Institute of Technology and Evaluation  
 \*5 School of Science, Kitasato University

\*6 Institute of Medical Science, University of Tokyo  
 \*7 RIKEN, Genomic Sciences Center  
 Following url is also available.  
 http://avermitilis.ls.kitasato-u.ac.jp.

# FEATURES

source  
 1. 300425  
 /organism="Streptomyces avermitilis MA-4680"  
 /mol\_type="genomic DNA"  
 /strain="MA-4680"  
 /db\_xref="taxon:227882"  
 /note="This strain is also named as strain: ATCC 31267,  
 NCIMB 12804 or NRRL 8165.-synonym: Streptomyces  
 avermectinus"  
 complement (382. .645)  
 /note="SAV258"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67967.1"  
 /db\_xref="GI:29603895"  
 /translation="MMNFALAPEVSVTDTEHTGTVLLDERTGRYWLNGTGALALRALI  
 DGTSPQGVASLQCFPDAGTAQVEQVDNLLHALRTAKVAAS"  
 complement (642. .2552)  
 /note="SAV259"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane-associated Zn-dependent  
 protease"  
 /protein\_id="BAC67968.1"  
 /db\_xref="GI:29603896"  
 /translation="MPAVRLPHLLITVEGTGMEFLVLPDPHAGAAALAMPTAGPPP  
 THVIEHSGREWIVCNNEEDVMVAVGHRVALLGRTTADVLRDVRDVSVRDL  
 DGLAATAGSFHVASVNEEDVQVTSACQIYATVAGVTLAARFQALADRIGAD  
 IDEQVLAALLSPFGPMPGLSRVCGSPPTGCLIEIGRDGTGRQWMTAPDPD  
 RPLGNDMIRQALAEAVATCTAGQEVVSADISGGLDSTLCFLADADRHDAHLTVHY  
 QALGRSSDRYWAQAEALPGARHLVPGQADPWYAPLAHAVHDSGLPLTFVSRRA  
 TTEHLARLAAAGSRHHQGGIGGDELFPSPMALHVRHPLAALPHLRALRSRRRW  
 PLTTTARTLASPSTQWLAACSNFFADRDHWGAGADWEPLPKLPWATPHAAATVRR  
 QLLAALATPQPLATLPVQHEMIRLTNRNGTATRENSHIGARTVTFHAPFTDRLVE  
 AILAVLRDLVALGQFKMLTALRSIVPAGLLSRTTKGDASAEIYAGLRRRRGLLK  
 LMDSHLYRMGFVDADVLRVLRGLHADTRPLMDLSDTLACELWRLTSLSPADTARATA  
 VVPSPRIVKEI"  
 3174. .3572  
 /note="SAV260"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="BAC67969.1"  
 /db\_xref="GI:29603897"  
 /translation="MGRAGPVAFLTDEQRNRYGTENQVDDQQLAGYFLLDRDRRA  
 MTCRGARSLQGVAGVLTGTVRFGLTFLDNPEDAPAEVYAVYAEQLGHPGSLVAGYGAER  
 TWMDHQIAIKDAYGYRDLKGDAAWKLKRDGD"  
 3604. .3834  
 /note="SAV261"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67970.1"  
 /db\_xref="GI:29603898"  
 /translation="MLGQQYVPLDLMIEGESQRMSEDRALRGLRMWVIAIEVPV  
 YLVWVEVSYLGGQQQFGRTEEDGKPRADTD"  
 4027. .4500  
 /note="SAV262"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67971.1"  
 /db\_xref="GI:29603899"  
 /translation="MOADAMAVVMELOGAAANRLWEGPARGRTFTLLTLAFAGG  
 AARARTAGTQVSGVAGFGQAAELLSRERVASKOTVAAREPLAQAAATAAAPLMRCS  
 APAVVATQQLNALADVNTARLEALEAFGRVNVATPPRLSWARRAVSSG"  
 complement (4811. .6445)  
 /note="IS6 family; completely same as SAV7545 and SAV75480





```

/transl_table=11
/product="unknown"
/protein_id="AA054114.1"
/db_xref="GI:21449382"
/translation="MVTASAVLCARAGCAPAGATATWRCSTACKQAAVRLRRKA
GSASHEDEQLREALALVARATILTEATPWSAQATALVGLARRVVELAAADAGV
SGS"
gene
complement (2129..3307)
/gene="asm35"
CDS
complement (2129..3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclomycin resistant protein"
/protein_id="AA054113.1"
/db_xref="GI:21449381"
/translation="MAVGKADSAWGVLPAALLIVALTGTITIDMVSPPDIARLGL
VGPAPVQLVTNLFGLGICGALCHGFLSDRFGRRPVLLGSAVMLGALCALPTIG
LILTARFVQIGASAAALVFVILIGDRYGAALRMVGLVNAAMSALMTCAPLGGAI
NRVAGRGNYLVACATLSLLAVFLFELPEPTRTRATITARSVLADYRLILGSRAYL
ASLTSLLPAAVLFIASPFYLTAAFGVLGTGEALHLLVVVASFVPSAFATGILP
LLGGPAGAVRWGLGAVVGVGLLGGFVAVIGSVLFCVGFACVCFVVFDRSV
FPLAGASATNGLRALLVTLTLAGALSTGGPFAPAAVMAGAAVAVVLLSALSFR
"
gene
complement (3450..4064)
/gene="asm34"
CDS
complement (3450..4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA054112.1"
/db_xref="GI:21449380"
/translation="MARNERAALDAAVEVLAGEGRLTFRVADQAGVPAGTA
SNYPAGAEILTAGERYVQRLDQAPIGETPGDRSRVTELMLHVLDRVSAFPG
FLALELRLEARNPELVETARVADLRFNAHHAASGLPGDGTVALHLLALNWL
ILRLTLPDLTQHORLDVLTALVDRLLAGQPEAGSPFDESQ"
4151..4723
/gene="asm33"
4151..4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AA054111.1"
/db_xref="GI:21449379"
/translation="MRKLVYFIASIDGVIAIGPGEHDFYTAEPWVHLRAEYPETT
PTRYPLVGLPTNRYDVTLMGRSGYQALDAGVGSFYAHLRQYVVSXSLPDADP
AVALVRADPVGLVRELKREDGLVLCGADLAGQLLPEIDELVVKYYPVVGSGTFL
FRAGFAPRFPDVRAFDHGGVVTYRPA"
complement (4854..5057)
/gene="asm32"
CDS
complement (4854..5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA054110.1"
/db_xref="GI:21449378"
/translation="MWITEQARIFARRSGRSNARRCAVRWFRQPRSATRPGFRS
SANADSRARSGRDPRTSINRSA"
5415..6461
/gene="asm31"
5415..6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AA054109.1"
/db_xref="GI:21449377"
/translation="MWIGYADGRRRRFRVWGVTEAFWETDEDKLAATLRDAEA
ANAASTRALYKAIQRVLLTADQVEVLARVVEAGLYAEHLDDLPPARGSRPVELR
RVVRDQQRKDHLEALNLRVLSIAKRYTGKMWLLDVIQEGNLGLVRAVEKDYAKG

```

```

gene
complement (6501..9518)
CDS
complement (6501..9518)
/gene="asm30"
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AA054108.1"
/db_xref="GI:21449376"
/translation="WVATGTRIPGPKPLVGNLLDVLTSDDLTDVDFELDRCHREHGG
IVALTPAQORQVFASHELVASDFSGWGAHVPALEQVRDFAGDGLTARGDEBNW
GKAHRLMFAFGFTAMDRHFMCLDIAEQMLVRRFRFPDHRDLDVADDMTRJLDTIA
LCAFGARFNSFYDRAPFVDMVRSVLEAGERAERLPGVQFPLVGNRQRYDDIATM
NRADGIVARAALPAGERPDDLRLRLMTCADPTGTERLSARNRVQLATFLTAGHET
TSGLLSFVHRLLAHEVLRKADAVDGLVDRVPAFEDLARLDYLGVLRETLRHP
TAPAFALPDEPEALGCGHAIGAGEPVLVMLFTLHRDPVAVWRDPDVPFERFAPERDE
IPACAMPPFGHARACIGRPFALQEAATLVLLVLRQFDLALADPHRLTIKQTLTKP
DSLVRARPRADPGATATVTVVPHQVPAFTHRGTPLVHYFGNSGSGEGLARTIAG
DGAARWATSVAPLDDAVRALPASGPVTVSSSYNGAPPDNAAHFVRLTQDGPDLGG
VDYLVLCGNLWSATYQVRVPTLIDMAAAGARLRLRGATDARADFFDGMWRWYEP
LWLLSACQVEVEIGPFRFVSDAAGLDGLASAVLENRELVRGPDAGSKRHE
LRPDGTSYRTGDLSPONHEDLVRRVAVARLGTAEARVTVVSSAPTGLVPSVGRAL
RVDELITRCVDSAPAGAVARLARCPFCPPRAELAAATGAILLELRFPSCAVD
LALALELLPAPTRLYSISAAEQRAEVALTVSVTGTSYLSRVPKRVGRVAGIAS
PPSPRPADNTVPVLLIAGTGIAFPGFLRARAALGGEPGALLFCGCGPELDDL
YAEFEALGDWLEVDVSRHPDGEVRHVQHRLQWRDRVRELVDAGARYLCCDATR
VGPAVEVLGRIGFGAGWLDLALRAGGYATDVE"
9664..10245
/gene="asm29"
9664..10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA054107.1"
/db_xref="GI:21449375"
/translation="MSAAVRREQLVAAALVRMKRKGIAAATTRAICAEADMPHGAFHY
CFRSKQELYTALLTGNTSLDAAPASIPDLPVAVNVNRTMLAAFTAFQADPRAELV
LGLVTLIAQREPELRPELRPAHQHREYVAKAVAHLDRLAEAEAGFDYSVDRLAELVVS
LDGVASSWFSHRDGLAKASLDQFVAMFAALTR"
10393..10857
/gene="asm28"
10393..10857
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA054106.1"
/db_xref="GI:21449374"
/translation="WTDTTTSHRTPVQNAALVAGVFLVLLGIAGFVPGLTNTYDGLF
ACHSDAALLGVFKVSALHNLVHLAFGRGTAAGLSKAPRGARFLIGGAIYAVLWL
YGVVVDHSTANFVPLKNAUNLHLGLAVGIGLGAALGRNAHYRAARSDD"
complement (10912..11937)
/gene="asm27"
CDS
complement (10912..11937)
/gene="asm27"
/codon_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="AA054105.1"
/db_xref="GI:21449373"
/translation="MPRELAQVWVVGASSGIGRVTALRFAAAGARVVCARNTRAL
DGLVEVRGAGRAVAVTADIADAENAPVADAVVERFGRVDTWNAAGIGVYGRVED
TPAGEFDRVMRVNLYLGHVHGAKAALPALRRAGGGVLLGVASVILGRSLAPLQAPYASK

```

Query Match 5.6%; Score 28; DB 1; Length 82746;  
 Best Local Similarity 100.0%; Pred. No. 0.017; Indels 0; Gaps 0;  
 Matches 28; Conservative 0; Mismatches 0;



QY 103 ACGTTCGTCGAGTTCTCGCGCAGCGCG 130  
|||||  
Db 22257 ACGTTCGTCGAGTTCTCGCGCAGCGCG 22230  
|||||

RESULT 27  
AX153707  
LOCUS AX153707 665 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 31 from Patent WO0140497.  
ACCESSION AX153707  
VERSION AX153707.1 GI:14535358  
KEYWORDS  
SOURCE Streptomyces ambofaciens  
ORGANISM Streptomyces ambofaciens  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Jeannin,P., Pernodet,J.L., Guerinneau,M., Simonet,P., Courtois,S.,  
Cappellano,C., Franco,F., Raynal,A., Ball,M., Sezonov,G.,  
Tuphile,K. and Frostegard,A.  
TITLE Method for obtaining nucleic acids from an environment sample,  
resulting nucleic acids and use in synthesis of novel compounds  
JOURNAL Patent: WO 0140497-A 31 07-JUN-2001;  
Aventis Pharma S.A. (FR)  
FEATURES  
source Location/Qualifiers  
1..665  
/organism="Streptomyces ambofaciens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1889"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 665;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGCGCGCTCAAGGCTTCGCGG 166  
|||||  
Db 389 CCGACGCGCGCTCAAGGCTTCGCGG 415  
|||||

RESULT 28  
AR362090  
LOCUS AR362090 15872 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 1 from patent US 6600029.  
ACCESSION AR362090  
VERSION AR362090.1 GI:33770246  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman,D.H., Williams,M.D. and Xue,Y.  
TITLE Metabolic engineering of polyhydroxyalkanoate monomer synthases  
JOURNAL Patent: US 6600029-A 1 29-JUL-2003;  
FEATURES  
source Location/Qualifiers  
1..15872  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTCGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 29  
AR362091  
LOCUS AR362091 15872 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 3 from patent US 6600029.

ACCESSION AR362091  
VERSION AR362091.1 GI:33770247  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman,D.H., Williams,M.D. and Xue,Y.  
TITLE Metabolic engineering of polyhydroxyalkanoate monomer synthases  
JOURNAL Patent: US 6600029-A 3 29-JUL-2003;  
FEATURES  
source Location/Qualifiers  
1..15872  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTCGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 30  
BD217171  
LOCUS BD217171 15872 bp DNA linear PAT 17-JUL-2003  
DEFINITION DNA encoding methymycin and pikromycin.  
ACCESSION BD217171  
VERSION BD217171.1 GI:33026941  
KEYWORDS JP 2002536959-A/1.  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 15872)  
AUTHORS Sherman,D.H., Liu,H.W., Xue,Y. and Zhao,L.  
TITLE DNA encoding methymycin and pikromycin  
JOURNAL Patent: JP 2002536959-A 1 05-NOV-2002;  
REGENTS OF THE UNIVERSITY OF MINNESOTA  
COMMENT OS Streptomyces venezuelae  
PN JP 2002536959-A/1  
PD 05-NOV-2002  
PF 25-JUN-1999 JP 2000557373  
PR 26-JUN-1998 US 03/105537  
PI DAVID H SHERMAN,HUNG WEN LIU,YONGQUAN XUE,LISHAN ZHAO PC  
C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/00,C12P7/ PC  
42,C12P19/62,  
PC C12N15/00,C12N5/00  
CC DNA encoding methymycin and pikromycin  
FH Key Location/Qualifiers  
FT source 1..15872  
FT /organism="Streptomyces venezuelae".  
FEATURES  
source Location/Qualifiers  
1..15872  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"

ORIGIN  
Query Match 5.4%; Score 27; DB 6; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCAGCGCG 131  
|||||  
Db 3721 GTTCGTCGAGTTCTCGCGCAGCGCG 3747  
|||||

RESULT 31  
AF082100  
LOCUS AF082100 28732 bp DNA linear BCT 27-OCT-1998

```

DEFINITION Streptomyces sp. MA6548 FK506 peptide synthetase (fkbp), FK506
oxidase (fkbo), and FK506 polyketide synthase (fkbb) genes,
complete cds.
ACCESSION AF082100 AF082099
VERSION AF082100.1 GI:3798623
SOURCE
ORGANISM Streptomyces sp. MA6548
Streptomyces sp. MA6548
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 28732)
Motamedi,H. and Shaifee,A.
The biosynthetic gene cluster for the macrolactone ring of the
immunosuppressant FK506
Eur. J. Biochem. 256 (3), 528-534 (1998)
JOURNAL 98451508
MEDLINE 9780228
FUBMED
REFERENCE 2 (bases 1 to 28732)
Motamedi,H.
Direct Submission
Submitted (04-AUG-1998) Molecular Pharmacology/Immunology, Merck
Research Laboratories, P.O.Box 2000 Bldg. RY80M-213, Rahway, NJ
07065, USA
FEATURES             Location/Qualifiers
     1..28732
     /organism="Streptomyces sp. MA6548"
     /mol_type="genomic DNA"
     /strain="MA6548"
     /db_xref="taxon:82632"
     complement(61..4575)
     /gene="fkbp"
     complement(61..4575)
     /gene="fkbp"
     /codon_start=1
     /transl_table=11
     /product="FK506 peptide synthetase"
     /protein_id="AAC68816.1"
     /db_xref="GI:3798625"
     /translation="WTSRGKRPGLTLLGSGPKRNSITNREIVAADRAEPLLASFAQR
TWLQQLDPSNAYNLPLVRITGLDALLERALTIVARHEALRTVFDADPEPQ
RVLPAPRLIHARADGQGAARLVHDSIAAPPDLSTGPLIRALLIRVGGDDHVLAVT
VHVAAADGWSFGLQLELAACHTALRDRSRAVLPPLRYADFAAMERRELTGAGLE
KRLAYMRDLRGAPVLDLPTDRPPVADAGLVEWRPPEALAAVRRIRDRAGTS
AFMTLLSAQLVLRHAGTRDVLVGTVPVANRTRTEFEVIGFVNTLALRGDLSGPS
FRELIDRCRAADAFADALPFENLIEVLAPDRDLSVNVVQVLLQVLRDASAPAL
PGVTAEPGTGHWFTFDLEPHVVEADGALTGELLYSRALFDESRIIRLLDEFTAVL
RAATAAGTRLSQLPGAEATPPEIPSNDRARELPVDTLPELLARIARTPDAAVTD
PCTSLYGELEDRRRLAHLLURAGTVTGDLVIGICADRGVDLVIGLIRKAGAAVTP
LDPEHPERTDFVLADARLTAVAHETYRHRPSGVPHVTDPDPLGVRIGTAPHLTL
GRDSLAYAIYSGTGKPKAVLMGVSAVNLILMQERTMGREPSTVQFVTATPDS
VOEIEFALGGTGLVTPDBEARDPPLGARWMDQOAITRIYAPTAVLRALVEHDPSD
OLSARHLCOQGEALADARLELCRPHRLRVHNYGPAESQLITGYTLPADPTWP
AAAPGRIDNTRIHLLDARLPVPDGMFGQLCVAGILGARYLARPELTARWYSGG
ATGERMNTIGLARARGGDDFLGRIDDQVKIRIGTEPGEVNEVAEDARVAHAA
VSVREDPRGEFLAAYVPTTAGPHGGLAASLREGDLRPLPAYLPVSAVMDALPRT
TSQKVDRLALPDPAEPGAPSKVAPRANRAERVVCRIFODLVDPQGVQIDDEFALGGH
SLIARVRSRIRAHQSGADVLRITLFDAPTAALARAADAGPAALPILSSPCAGPAP
ITAAQOMLHSHGSLAAPSITYVAPVGFRLGRLDRLDALTAALTRITARHEPLRGER
GREQIVGPPAAVCAEVPEVSGDVGAARAVARAELARPDDL VAGLIRLAVLLPGAED
HVLLMLHAGDWSFDLLVRELGSDDLPSVPDVARWERVFAVVAARNDRAYW
RRQFGACAPELPTVRPRTAAPDGRFLMTDDATVAARARVADARNATWHTFVGA
FALVAAEADDTDLVATPFAADRGVAGTDHLIGFFAKVIALRIDVSGSGTGTPSPFAEY
LRVHTMTGAHAGSVPSYAMRAADPALPPTPVSFQLISALSTELRLPLGLTEPPPV
VTVTDEINGELSVNLFDGGHTVSGAVVHDAALLDRVYDILLTRVESTLRRAAGDLT
VRTVQVQSGE"
     complement(4572..5570)
     /gene="fkbo"
     complement(4572..5570)
     /gene="fkbo"
     /codon_start=1
     /transl_table=11
     /product="FK506 oxidase"
     /protein_id="AAC68817.1"

```

```

/db_xref="GI:3798626"
/translation="MSAPVAAPCYCRFEKIGSPDLGDETVLGVIEHGTGHTGASLVDG
VPHALTHTTRDEEAFAEVMHQAOPPVESGVTDDGIAWARTDEYDFGVGRVPSGRYADA
VSALYTRISGLTRSLGHPILLARTWNVVSGINSANADGLEVYRDFCVGRVPSGAGGID
PASPAAATGIGTHGGIACAFIAARGVRNIENPAVLTAHHYPAAYGPPSPGFARAT
WLGPPDRGLFVSTAGITAGLHETHHGDVTVQCEVALDNTARVVUGAENLRHGGIRGH
LLADVHLKVIVRRREDLAARRVRCARLSTQATTAIVLVHTDIARGDLLLEGVVR"
5798..28528
/gene="fkbb"
5798..28528
/gene="fkbbB"
/function="FK506 polyketide backbone synthesis"
/notes="initiation and first four elongation steps"
/codon_start=1
/transl_table=11
/product="FK506 polyketide synthase"
/protein_id="AAC68815.1"
/db_xref="GI:3798624"
/translation="MAAESIRPLHLLRVNAERHGRVATDSRRVTVYQLQIRTC
RIAGHLAASVGRGDRVRLLLGNRIETIEVYLAARAAAVAPLNPDAAGDLAHLFT
DSGATVLTDDLHDQVRRVGTDATVYLAGRPEPCVAYEDLACTEPRPDRDLGLD
EPWMLVTSGTIGRPGKVSQAORSGLMSATCDVPSWELTDELLMAPLPHSLGHH
LCILAVLVGASARILGSFVARDVLDALAEHPCVLVGVPTMYLLGAVSGEPQTRA
LRVALVAGSTSPASLTAFGEAAFGVPLIDTYGCTETTGSLTANTLTTRVPGSGLPV
PGLSRFVDPVTVGRGDELUGELVAGSPSLMGLYHAQTEATAVLADGWRTGDLAR
QAETHVITIGRVKELIIRGENIHPREISVQAQVPGVRDAAAGAPHPVIGTIPVL
YVPEGVPADALAEBCRQLAYFKVPELHVTVTPRTASVKVRGELAGLEAPLV
GTSGEALCELWMERELPDTVSQVAVVTRAVGIGPAEVPDADQALRWDEALVRQ
AABPGSVLVDTDITDNDANDEDELVAALGEPVALSGVAYVPLVRADIT
TPIPASQWALRPPASGTIRDLAVAPSDAPPPLAAGEVRVDRAAGLNFRLVLTALGT
YFGDAMGGEAAGIVTEVPGVDLVLGAHVGLVQDAFRSVVADRELVARIPGWS
FVAAASVPIFATAYGLVDAGQVRPGKVLHAATGVGSAQAQIARHMGADVYATA
SPAKHLLADDFADHANSRTAFADAPFVVDVNLNSLTGEFLASVELLAGGRF
IEMGKTDIRHDVQPPFDLSVDPALPVLITGEGADAVAGLRDLETGRTHVSSAAEA
FGWMSOARNTKTVLTPRPLGPDPVLITGEGADAVAGLRDLETGRTHVSSAAEA
BAERGESGNASTLAHVLLGDPACADAYARSASEGLPIVTVHGLPGTEQAALTCELI
EALIEGGSVYVTVRSGAGARAEMAGANTPILSGLAGTGRGAPGAGPWGSLAA
ADARENVLLDVRDVSATVLGLPGARCAFDRTIFRENGLSLTITVEFANTVARTGL
RVPASTDFHTPEPSFAAHLGSAVAAAPAAKALGSEPAIVGMACRLPGGVTSPEGL
WRLVASGTDAITPEPDRGMDVDELDPDPAVGSRTHHGFTEATFTGDAAPFGIS
PNEALMDQORLVLETSEWAEFHAGIVPDTLRESLDTGVFMGAGHQYGAGDLGGIG
VTATQTSVLRSYFFYGLQGPATVDTSCSSSLVALHQAALRSGCSLALAGVGT
VMAITPGEVSERQSGUSPDGRCKAFADAAGTGFAEGVGLVVERQSDAERNGHTVL
AVERAVNQDGSANGLSAPNGVAQQRVIRQALANAGLUSGADVDVAZHAGTGTVLGDP
TEAQAVALSYQEREVEPILLGSVKNVGHTOQAAGVAGVIMWAMRHGLLPTLHVD
EPSHVDVMSAGAVEVETREPSPSGARRAGVSFGVSNHVLLEGVAESSVRDG
GSSGLVPLVPSARTESLALOVERLGEVYRGADAAVADGLVRCRAVDFRVAVLLE
SAYAGAVEGARTVFPFGQSQWGMGRELWGASEVFAARMRECAAVLEPHTGWDLL
DVLGEAVVVDVRVLEQPASWAVASLAALWQAHGVVPDAVVGHSGEIAAACAVALS
LEDAARVALRSQAIAARLAGRGAMASIAVPAASAVETVEGVWIAARNGPSTVAGDP
AAVERVLARYEAGVYRRIADYASHTPHVEAIEAQLADALEGTTSTFSPFWMSV
DSGWTEPFGAYWYRNLQPVAMDVASELDSGLFIECSAHFVLLPALLDQERTVASL
RTDGDMDRFLAALAAQATQGVADVMTTLEPAPHRVLDLPTYPDHRXYLQLPAPV
GAGIGHPLSSAAILPGSDGLVTKLSLATHPWLADHTVOEVLPTGLTFLVRA
GDEVCDTIDELIETHSLPASGAVDLTVTVQDDTGRHPVSVHAPAGQDWTWTRH
ATDGLRGTEDAPTNSFPMPPTGAQSPASLDEFERLAGAGTGPSPAPQARAWRA
GDTVYVALDGAADVDFGHPHLLDALHAGRLDAGDELELPEPSVGRVRLHAGAA
AVRAHSGPAGVAVQVADLDGRPVSDALVRPSVAPSGPNLHGLLEWPAEABH
DGADLPGYTLITATHPDDDDPTPNHPTTRTQTVTLTALQHLTTNTLI
VHTTDPGAAVTGLTRAQNEHPCRJHLETHPHPTPLTLTQTLTQHLPHLPLTSLR
LHTPHLPTTHHTTTTTTPTPLNPHALLITGSGTLAGILARHLNHLPLYLSR
TPPTPTPTTHPCDLTDTQITQALTHIPQTLTGFHTAATHDATTNLNTPQHLTT
TQPKADAAWHHHTQPLTHFVLYSSAAATLGSQGANAYANAALFDALATHRT
QLOPATTAIWGMHTTTLTLSOLTDTQQRVDGRFPLTEAGNEHFLDASLADTAPFV
VAAIPAEARREVERTGTGDSGRDILLAVCVATAAVALGHADASEIGAAAFKDLG
VADLSGIRLRSLAEATGRLSATVDFHPTPDALAELELGDSDSQGAAAAAA
PGTVFAAAGDEPLAIVAMACRMPEGVDPEDLMSLVESGSDATTEPTDRGNDLAA
LYDPDPAIGKVSVRHGFLEGAADFDFEGISPREALMDPQRLLYEVESWEAFER
AGILPASVRGSDAGVFMGAFTQYGAGVDLGGFATGTPTSVLSYFVLEGPVS
TVDTACSSSLVALHQAARSLRSGECSLALVGGTVMTATTFVFSRQRLGAPDGRAK
AFADTADCTSPAEGAGVILVELSDATRLGHFVLAVRGSAVNSDGSANGLSAPNGPA
QQRVTERALDDGLAPGLDIAVEAGTGTGRLDGDIPEAOALEAAYGLDRLHPLLIQSLK
SNFGHTQAAAGVAGVIMWAMRHGVLPRTLHVDFSPSHVDVGGVGLLRRLRNEPFWT

```

GRVRRGVSFGISGAGAHVULRAGPAPAPLPAADVPEDVPMVPSARTPDGVRD  
 AGRLTALPAPAAI GHSIATRTAMHRAVVPADAEAFARGEVEVGVVGTADVTDT  
 RAVVFPQGSQMDGMAELIATPVPARKIGCEALAPYTGWLLDVIARSGAPE  
 LDRVDVVPQASFAVMVALAEHAGVAPAAVGHGSEVAAACVAGVLITDAAKV  
 ALRSIVATERAGGGMVSPADFDAAWAGRIEVAANGVAGVIVVAGAAVEEIL  
 AYPHARRIADVASHAHVIEHGRVFDALADLTGAPBIPFESTVDEAWLDRPDA  
 AYMNVRCVPRGAAARAEIHLGRVFBASPHVLTALADTLAGHNTAVGTGLR  
 RGCGARRFRSLAELWVRGVFVSWPFGELRGVPLTPVPRDRYVDAEPAGTSGHP  
 LLSGLVERAGEGAFALSVKQFVWPLADHVDGRIIIPGSAIVELLAEAGLGT  
 TIAELTAVGVVDEEDTHIQTAVTEVSGRAVRHRSRIGAGTVPVGTGALSAT  
 RDEPEWMPADAPDVLGTDTLPISYGPAFAMTAMWAGQRAVASVRLRGQDIT  
 RYGLHVLDDAAVHALGTIADPERRLAFNSGVRHARAATVDEVLLEADGDTI  
 RILATDEHGLVGLTVARABEPTALFEVAVWSVPASVPGWTHLADVPBGDVI  
 EGADGEDVESDPPVVLVAVEPDPSGVRARELGRDLATVTLWLAAPWAS  
 RIVVVTGIDPQAQAGLVRTAGTEHPGRVGRLEADEITPATVEAGLAAGDRSHVR  
 TGTVRAARRVTAAGSPSGTGTGTGTGGLRGLLVHMLGIVDDAVATVRPQ  
 GGRPDGAPDRVRYVADVTRDELAALVDGVAERLCAVVMAGI VDDAVATVRPQ  
 QMDAVLRKADVAMQWHLTRDLAEALFVLYSSIATFGGAGQANYATGNALFDLAR  
 HRDQGLPAYSLAWGLDEADMGRLAATDTRIARGGMIPTMAEQGLELFDALHS  
 DRAALVRLDLAAVASDQVPIRLDLVNRARAAAPTAAARDMLELVRTSAALVIGH  
 RDAHAEPTRAFKEVGDLSLGVBLNRNLADATGTLTPATLVDHPQAALAAHDEL  
 AGARTATRTTAPARRHDEPLAIVMACRUPGGVSPDDLWRLLSGSGDITAFPAD  
 RGDVVDALYDPPDHSGTSTVHGGFLTGAGDFGFISSPREALMDPQVRVLET  
 SWEALHAGIDPHTLGTGTGTFVGMGYFYGSGADRGDFGATSTQTSVLSGRISFY  
 CLEGPATVDTACSSILVHAGOSLRTGCSIALVGGVTVMASPSGVDFSOQGL  
 APDGRCKAPAAAGTGAAGSGLVWVRELSDAERHGRVLAIVRGSVAVNODGANGL  
 SAPNGSOVRVIRLANAGLQSDVDAIEAHGTGRLGDPIEATLALATYQDRATP  
 LLLGSLKSNIGTHQAAGVAGI IKNVLMWRHGTIPTLHUHPTDPSHVDNSAGTVELLT  
 DARTWPTGTHRAAVSSFGVSGTNVHIIIESHPRTAPDTPGSSVTHVPLLSARTP  
 EALDEHTRVAFDAEGEDERAVASALTTRTAFTHRAALI GTDLITGAEPRRLW  
 LFSGGSORPCMGDLAAAYDVARTRDVLDAQVPAAGLDVHDYTAQPAVFAQLVA  
 LSAQDANGVRDVLVGHISIGELAAVAVGWSLDDATELVASARALMQLPQGMAMA  
 AVSASERDALPCLCEVEIATHEPKLEMAAGADCTPEYVVRQVDTVRFAEQVAYDGA  
 ARMEPLDEFDAERLTYHPEKLEMAAGADCTPEYVVRQVDTVRFAEQVAYDGA  
 ALLETPGPRNRLARLDVGIPLVHGEDEARSAMTALARLHGGVAVDVEVGAAPTDL  
 HLPTVPRRTVRLGSAAGAAAPAGQLPVAPVLTAAVMVPGTGLDLVLTGRVEATDP  
 LATRVHPCGAARVLDLRSRATKRAAPWRSPWTPRSALPQSGGLALSVTVAAPG  
 EDGRANAVHARPGTGEWTHATGILAVARVASAVPTPSPWPPATARPFDVGBLAD  
 RIARAGHTDGPALPRPAWADDAAVHAEVALADQHADAERYGLHPALLGALALAG  
 EGADLPFADVYAEVATGATTVAVTATGTHLADETSFVAIVAGVRRRPLTVEGAV  
 PGLLRPLDIAELAEIATTTGGLDPPVVDVILPAHGTGGGPGIETREDLGVILAA  
 LRSFLTDREYADAVLAHVTPGLAPAAAGVRLTAQAEHPRIIVVDAEPDPAELLA  
 AAGLGEPOVTVREGRVARRLTAPVPSGDAPELDPDGTVLITGSGTLAGIVARHIV  
 GHYVRLLMLSGRTASDVGAEVATACDIADRELASALAGIDPAHPLTAVHTTA  
 AVIDDVLTALPERDVLPRKADGAWHLHEITKDDIQAALCSTFRPEPCASDREA  
 TTPRTLLDELAEORRAGGLPALSVANGWEPSGLTQAGDRMRDGVYALSTERGL  
 ALDLAARSPALVAADPGGLGSSALLRAPHRRPRAAGSALADRIAGLSAADAE  
 KAALAVRCAAAVLGHEGPAREATATFELGVDLSLTVRLNFAETATDVRLPATA  
 VDFPTPOAAKLTALTCVRPARATKAVGHDPEPLAIVGMACRLPGVVASPDIL  
 WRLLSGGGGICITAFPADRGWDVADLDPDPEHSGTSTVRHGGFLTCAGDFDAGFGIS  
 PREALAMDPOQRVLETSWEALQAGIVFTGLRGSTGVFMGAFSGYGLGTDLGGG  
 ATGTQTSVLSGRISYFGLGEPVATVDTACSSILVHAGOSLRTGBCSILALGVGT  
 VNASPGGVFEFSQORGLAPGRCKFAEAAADGAFAGSGVLTVDSEARHGRVLI

Query Match 5.4%; Score 27; DB 1; Length 28732;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 280 GTCAACTCCGACGCGCGTCCGACGCG 306  
 Db 14927 GTCAACTCCGACGCGCGTCCGACGCG 14953  
 RESULT 32  
 SPAS580915 74787 bp DNA linear BCT 10-DEC-2003  
 Streptomyces parvulus Tu4055 cluster for borrelidin biosynthesis.  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 AJ580915.1 GI:39725416  
 2-hydroxyhept-2,4-diene-1,7-dioate isomerase; actolactate synthase;  
 aminotransferase; borA1 gene; borA2 gene; borA3 gene; borA4 gene;  
 borA5 gene; borA6 gene; borB gene; borC gene; borD gene; borE gene;  
 borF gene; borG gene; borH gene; borI gene; borJ gene; borK gene;  
 borL gene; borM gene; borN gene; borO gene; borrelidin polyketide  
 synthase, type I; cytochrome P450; dehydratase; dehydrogenase;  
 ketoreductase; muconate cyclo-isomerase/lactamase; ORFB3;  
 oxidoreductase; thioesterase type II; Threonyl-tRNA synthetase.  
 Streptomyces parvulus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.  
 Olano, C., Wilkinson, B., Moss, S., Brana, A.F., Mendez, C.,  
 Leadlay, P.F. and Salas, J.A.  
 Evidence from engineered gene fusions for the repeated use of a  
 module in a modular polyketide synthase  
 Chem. Commun. 22, 2780-2782 (2003)  
 Olano, C., Wilkinson, B., Sanchez, C., Moss, S., Sheridan, R., Math, V.,  
 Weston, A.J., Brana, A.F., Martin, C.J., Oliynyk, M., Mendez, C.,  
 Leadlay, P.F. and Salas, J.A.  
 Biosynthesis of the angiotensin inhibitor borrelidin by  
 Streptomyces parvulus s Tu4055: cluster analysis and assignment of  
 functions  
 Chem. Biol.  
 3 (bases 1 to 74787)  
 Olano, C.  
 Direct Submission  
 Submitted (19-AUG-2003) Olano C., Biologia Funcional and IUOPA,  
 Universidad de Oviedo, Julian Claveria s/n, 33006 Oviedo, SPAIN  
 Location/Qualifiers  
 1..74787  
 /organism="Streptomyces parvulus"  
 /mol\_type="genomic DNA"  
 /strain="Tu4055"  
 /db\_xref="taxon:146923"  
 /note="Streptomyces parvulus"  
 <2..313  
 /note="ORFB1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAE45653.1"  
 /db\_xref="GI:39725417"  
 /translation="IPRGAVDVLRAQDRWPHSAAPRRHGLLNANWGANWVATVDFRY  
 ASRTYDDQADQVDIAIHDAAGLVNAGAGFDILAVALILFVRLTAQAHAALAGATPTPT  
 H"  
 501..3107  
 /note="ORFB2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAE45654.1"  
 /db\_xref="GI:39725418"  
 /translation="MEAFLLAAESVLLRRDQSVYVTPGSEDPGPAPRALRLRLAEELL  
 GRHVASPLHVALELSEELAAHVRVLVPSQCAHLVCLRCWDSGLDYGACPLCHRR  
 EALYDVRVAFLLQPEQPCVLCGEARTVLVPSQCAHLVCLRCWDSGLDYGACPLCHRR  
 IDGDDEFLRPRVAVGAARATVPGLRLRLGLTMDTADATVADALLARRPLSPQDRD  
 RLITLPLTPAGGDLPPQIPVRETKALVGLVRRAPSPERPALELLARLTATDVL  
 RLAVLSSGAGLVTPTFTNPRSLRLVLLAVDLPAPYLVEDMLRHPTANKRAE  
 WLPFPEGHTRHRAALATAVRATPLDPTAFGAALLTTAAHPDAVPRDGTVRPAT  
 WAGRLSQMAEGDAARAALAGERPGLVRLDLVLRHDEALVPELEKALRHGLPK  
 VGGPILLALGALRTETDGTTRVFFPRGDVTRALSVPERRPALPAGVSEVALL  
 EGGLERRFAAGRLVSLVDAGLTLDTLPTERTAKALVTVGRGSVQALPAGVSEVALL  
 FLWTFEFGNRTDLISLVAFFDAEWTFITGLCDTNLVHGDAAIHSGDLTSPAPRGA  
 TVRFDLGLGSKVCLPMI VDLARRALWTDLHPSAGFQSI GSHGGGLAAGVAGDLM  
 QQTSGRATLWDLAVLRAALSPVAVVSRPEPAVLRYRRAAESEAFVAVASH  
 KDAERLHATDPSAAGLAAGRGVFLATVHGDVPRPGASGTSYRLFPAGDASPTILT  
 RVTAGDILLAEGL"  
 complement (3172..3810)  
 /note="ORFB3"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAE45655.1"

```

/db_xref="GI:39725419"
/translation="MAEQIAGIEIPDSAPAREATDLIRDTTPPLIFHRRVVLFGSL
QAAALGIRPDPELLYIAALPHDTGLVPPYRGDDQRPMDGDAQAHAFLLAHGIPRADA
DTWATAVALHTTEVPYRWAPEIAATAGVETDVLGRNLGRLTRAQIDAVTAHAPFD
FKQILRAHTEGFERPATTFGTGVNADVLEHFAFGRRTFDFEVIENSAPWE"
complement (3935..4924)
/notes="ORFB4"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45656.1"
/db_xref="GI:39725420"
/translation="MTARHSVGIILVFDGMKMLDLSGPAEVAEAFNRFQARVRLGVVS
PGAPVRSIGLIVPAEADARGAPDPTLVVGGDALPGSPVDPRLIDAAKALAAAPAG
RVASVCTGAPVLGAGLLEGRATTHOHTALARECPSTRVEPDAIFVKDGTATYSA
GVTAGDIALALLEEDHGLDARVARSLVYLORAGGQSASLRGAPRTPVLRLQ
VQDAQADPAADHSLAALAAARVRSPLHTRMFAELDVTVPKYELIRPDIKALLD
SGHNATEAALSGFPSEYSLRFAFARHLGLSPTRYRQRFATTPVDAGPRPDG"
5123..5953
/notes="ORFB5"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45657.1"
/db_xref="GI:39725421"
/translation="MGTVTTSQGTSTIFYKDMGPRDAPPIVFHHGWPLTADMDNOMLF
FISHGVYTAHDERGHRGSGQSTGHMDTYAADVAALTEALDLRDVAHGHSTGGGE
VARYARAEGRVAKAVLGAVPPVNMVQKSDANPGGTPIEVDFGFRFALANAAQFYID
VPSGPFYRNGRAKVSQGLIDNMVQSGAANAHYECIKAPSETDFTEDLKAIDVP
VLVAHGTDDQVVPYADSAPLSVKLLKNGTLKSYEGLPHGMLSTHPEVNVNPDLLDFVRS
"
complement (5961..6518)
/notes="ORFB6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45658.1"
/db_xref="GI:39725422"
/translation="MGVMTGPACRBERDEGHVTVQAPVATDERRVVDKOTPGAYKAF
VAAASVRRMAAAGLDRLVELNVRVSQLNAQACLSLHTAAALRAGETORLAVL
PAWRDTLFTARERAAALAAEATRPDAAAQAAAYQAQARGVLSDDSEVSATVIAISI
NAFNRVSLVSKHPVRGAAPAPVSPA"
6567..7538
/notes="ORFB7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE45659.1"
/db_xref="GI:39725423"
/translations="MGNTEPRPAEMRCGALEDEVPAAGVVELTARDVPLGPPRAMTVR
RTLPPQARTLIGAWCFADHYGDDVAASGMDVAPHPHIGLOTVSLFSGEIERDRL
GTHAFVRPGLNLMTGGFGIAHSEVSTPDTVLHGVLVVALPEERHDTGRDQHPAP
APVAFDGTARVPLGSLAGTSVSTFTPLLGALLVLPGGTATLDVDPGFHGLVD
SGDVRVEGAVRPAELGYAPAGRTLTLTINESAAPRLILGGLPPPEETIMWNFTIG
RSHDEIVRAREDMKGRDFGEVHGDCAPLPAPELNPAPLKPRRRR"
complement (7603..8397)
/genes="borB"
complement (7603..8397)
/genes="borB"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="thioesterase type II"
/protein_id="CAE45660.1"
/db_xref="GI:39725424"
/translation="MTCTNTHSDVWIRQYRPAHTAPQLICLPHAGGSAFTYHPVAA
LAPCDVLAVQYQCORRAEKPLEIDELANQLFPVLRARVOPVAFHSGMGATLA
FELARPESAGISEALLVSARPAQSPQRTGGTVHLLSDRELVAELTLDGTAEVPH
DEELVRMALPAIRDYRAATRYRPPKLRCPHALTGDDDDPMTVPVPEARWSEHTD
GPFTLDTFAGGHFYLLEHREDAILGIIAHLRTCSRAGDRSGSLTRE"
complement (8397..9194)
/genes="borC"
complement (8397..9194)

```

```

/genes="borC"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative ketoreductase"
/protein_id="CAE45661.1"
/db_xref="GI:39725425"
/translation="MSLELTDVVMTGAGSIGRAAARLLVGHGARVVLVGRTEGAL
TETTAGLPSHHLVVPCDDKQVADCVARAVRFRGLDDGFNNAAGTGFSGFPLHQD
TADNFDRTATNLRGVWMSMRQIEAMLTAGGAIUNCASVAGHAGOSPLYSATKH
EIVAGSLVALQYAGDGIKRVNVSPGSTTTPMLRSIYADPSALAQARAPRLRLKGC
EVANAVVWLSPLAAYVTGTVGVDGTPAGSALPRITNATPEGQHR"
complement (9244..9996)
/genes="borD"
complement (9244..9996)
/genes="borD"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative ketoreductase"
/protein_id="CAE45662.1"
/db_xref="GI:39725426"
/translation="MTARHDLVALVTCAGSGICAEVARGIAARGLRVLLDKDAEAVHR
VADGLDLRALDPLVADVTDPHALASAVDSLAPQHRPGLVNGVGGDTTARSVELITE
ADLQEAFTNLASVFTMTRLCPVMAVAGWRVNLASVAGRTYTFPSNAAVVAAGAG
VIGFTKQCAVELAPHGVTIVNAHAGVIGTERIRRAWEDKPPQWTADRVSHIPAGRFGS
VAAAGMVCHLGEDAGYTTGTVDVNGGLHI"
complement (9993..11165)
/genes="borE"
complement (9993..11165)
/genes="borE"
/function="borrelidin biosynthesis"
/codon_start=1
/transl_table=11
/product="putative muconate cyclo-isomerase/lactamase"
/protein_id="CAE45663.1"
/db_xref="GI:39725427"
/translation="MIRRVRLHTVVPMAAFHATRRSRASLILVEIELACTRGMG
EGAPRVYVGETLDGAVRAVQCDPCELAERIEWRDFESAVASIAQLPLTGLVDGSSA
AAVEALDLDVCRHEAPLADVLRLVAPPARRDRGPTSVSLVHLSRDVATVLDAL
TFPALAAKHVKIKVADPAGAVDLRTAQDRLPADTRVLDVNGANTAEAEKAVGEL
DGVGVEEPLPPRSPWELGRRRLATGLPVLMDSCCTGPADLHAAATSGRAASHINVRLS
KCGFLAAARLARLADDELGVGQQLGVHVAEGLPAAWAGRTLATAMDWLQWTEAGRADE
WFEVPLTTPAFTVDRSLHVRVEPLTGPCTGTEELLRTRCAATWESGGMRRT"
complement (11162..11980)
/genes="borF"

```

Query Match 5.4%; Score 27; DB 1; Length 74787;  
Best Local Similarity 100.0%; Pred. No. 0.052;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	136	GCCCCGACGCGCGTCGACGCGGTC 162
Db	34938	GCCCCGACGCGCGTCGACGCGGTC 34964

RESULT 33  
AB070940  
LOCUS AB070940 104326 bp DNA linear BCT 22-JAN-2002  
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.  
ACCESSION AB070940  
VERSION AB070940.1 GI:15023967  
KEYWORDS Streptomyces avermitilis  
SOURCE Streptomyces avermitilis  
ORGANISM Streptomyces avermitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,  
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,  
Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.  
TITLE Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary

metabolites  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
 MEDLINE 21477403  
 PUBMED 11572948  
 REFERENCE 2 (bases 1 to 104326)  
 AUTHORS Ikeda, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan  
 (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)

FEATURES  
 Location/Qualifiers  
 1..104326  
 /organism="Streptomyces avermitilis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:33903"  
 complement(1..1134)  
 /gene="mcma"  
 complement(1..1134)  
 /gene="mcma"  
 /note="methylnalomyl-CoA mutase"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="BAB69185.1"  
 /db\_xref="GI:15823968"  
 /translation="MTGISTADANARFQLVSGCTRAYVVFVFLPFRAGLSDAPAA  
 RLGGRRGVAVSIDIOMRVFLGDRVRVPSLVADPAAPLLVLIQLVAEHEGPAQ  
 RIAGAVRGVDFEFLARGVFPFLPSRLMNVAYCOAELPFWQCISVCGHHAGT  
 GADPALVAFTVAGGIEYLRAAVTAGLVRTVAPRLTLCLVGGTTGRENRAKLRT  
 VMTANERFGSGRTAPRLVTRPDGLATLACTEDIEAGATLVDRIERLGGTLTVL  
 EHLGAGQLGRPHGDRPAPRPHPHQDSLDIQAOTERLAKLAWRVQPAVDEALLR  
 LRKAQGDNDVLYPMREALAAATVGEVCDALRSWTGTSAT"  
 complement(1131..1886)  
 /gene="ppta"  
 complement(1131..1886)  
 /gene="ppta"  
 /codon\_start=1  
 /transl\_table=1  
 /product="phosphopanthetheinyl transferase"  
 /protein\_id="BAB69186.1"  
 /db\_xref="GI:15823969"  
 /translation="MRAHVFDVDPVSSLHPREALVQDAVSRREFATGRHCARRA  
 LASLGVPPGLATPGAPWPAGVCGALTHCAGYAAAVARAPGIRVIGDAEPDP  
 LPEGVLEVALPARQVKEHTARHGVTGKLLFSKSYKAMFPAGRPLGFEEA  
 LIDLDPSTFRVIFPTALURAEPELRDLHGRWLVRHGLIIVTAVVVDGPRDTPAP  
 SPYDRTKSAFTGSRADRPGRPTDRPTA"  
 complement(1910..2875)  
 /gene="simX4 homolog"  
 complement(1910..2875)  
 /gene="simX4 homolog"  
 /codon\_start=1  
 /transl\_table=1  
 /product="simX4 homolog"  
 /protein\_id="BAB69187.1"  
 /db\_xref="GI:15823970"  
 /translation="MSQDRHDSSTAALGGPSHGCRLLAVSDIHRHQRNDIVQDLHP  
 GCADWLI VAGDVSAVDITWALSILRLDRFAKVIWAPGNHELWTFAPGDEURGERK  
 YRLVDLCRALGVTPEDPYVWHGAGGVAVAPFLIYDYSFLPFCAMSKSHALRIA  
 DDAQVCTDEFLFHPDPYPSREAWCRAVRATRLADLPDLPVTVNHFVLRPT  
 RVLHPEFLAWCCTTATADWHLYRAAAVVGHLHTPTIRCDVDPHHEVSLGYPREW  
 QRYGRPDWVQILTPAPAPAPVPPVPPAPPGHRTASPADRP"  
 3213..3971  
 /gene="olmC"  
 3213..3971  
 /gene="olmC"  
 /codon\_start=1  
 /transl\_table=1  
 /product="thioesterase"  
 /protein\_id="BAB69188.1"  
 /db\_xref="GI:15823971"  
 /translation="MSANRVDGNLWVRQHPAPTAATKVLCLPHAGGSASYFLPVSKA  
 LAFGVDLAIQYPRQRDRREKCIDNIAELADAVTAVLEPLADRPLVLFCHSGATILA

FEVALRLRRKGIPTVLFASGRRAPSRHRDESIVHLRGDDEIIAELKTLGSTRSEVLGD  
 BEILRMVLPAIRSDYKAAETRYAGCARLTAPIHAHFGVEDPRVSLIDEAQAWADHTTG  
 HEELRSHPGGHFYLNDQAPRVLEEIARVSSSVAS"

gene  
 CDS

4643..7438  
 /gene="olmR1"  
 4643..7438  
 /gene="olmR1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="luxR-family transcriptional regulatory protein"  
 /protein\_id="BAB69189.1"  
 /db\_xref="GI:15823972"  
 /translation="MLIGRQGEIDRLRAAYSSVRSTGRSAVLVSAACGKSEILED  
 LIVRVESLGLIALRATAIERNTPLSVRLVNGAPFPDDLVRPFTMTSDAAARMIS  
 TRGRDKARRDAWYELMRTCTCEVRALAQAPLVAVDDIHGGDADSLQLLHARHC  
 RNVPVLLFTQSLAGPADRTFGAELLQPNLEHRLPRFCQDIALMLEAQCRADAG  
 QCTVERIFAVSGNPILLRALLEDDARLEPDDTVDQDVQPRAGQAYGRAVLACLRGG  
 EKAHEVAAGLAVLGAFTHELSLRLDISAAALTENLRALRAAGVAVGVSFQHPAAKA  
 AVLDSEPSHRLRLHQRSAALHSAAPAAAEHLHEAGQADGAWAVLQQAEOQA  
 LLADEARRAVSYLELAHAASADPRGRTEIKIRIGSVTRVSAADAEHRVDDALRGFGA  
 GLLTPSDTGGLAALLHGRLEQARSVLEQGRPGDGDADRTAARQLAEIRDSVLNR  
 GWACAGTRAWHENPAGRPSTGRAGTGLPAMLSSGSHAAALQRIACTEGLELTHLTD  
 ATFDLVNAVMTLLGALGAVERALHWCDFTLREAKRDAAGWYFAVFAEAGNALRGN  
 LAEERDARRCFVILPERHRSALGESVSVLVTVQIARGKIEDAARQLRPMHSDLPD  
 TTHMLDYLRAGLYMATQYHSALRDPHEVGLRAERWGTDPALLPWRDTDAELLR  
 LIGERQGLRLATEQLAMVODGAPRIKGISRLQGLASEPAARLGLLNRVAVELSRADD  
 RLELSRTPLDLADAHRETRQTRQAAMVLRARQLAKSCGAQPPDRIGSPDAAGSGE  
 LIGEPERNADNGHLSSESERVAGLAASGYTNRISVRLYITMSTVEQHLTRYVRKLK  
 INGREELLAARAGVQV"  
 7658..10486  
 /gene="olmR11"  
 7658..10486  
 /gene="olmR11"  
 /codon\_start=1  
 /transl\_table=11  
 /product="luxR-family transcriptional regulatory protein"  
 /protein\_id="BAB69190.1"  
 /db\_xref="GI:15823973"  
 /translation="MLASAGQGGVVLVNAAGPGVGTALLDLAFLRKATESGARVCAA  
 AGSSAETAEFGVVQQLFPPTGTLGAGFLARAIGSRDAGPGLDADRLFDALCQELRR  
 RPTVIAVDDAQLADAAASLRLRYLVRRLTAPVMVLTPTGPDALPFPFQAEILLHP  
 QCRRLGLPLSRDGVTVQVVSAAQSDRVRPLAADFHAISGNNPILVHALLHRAHRH  
 LSPFVPTVPVSSSAASASASASASASASASASASASASAFKAAALAWAYRDEVLV  
 DVACGVAGSATPARVALCDRGADAVDVMVLTALDTAGLGGTFRSALVKGALLLE  
 HDVETRAELRRERAEALHGDGAPVVAARHLLAAPVAPFWAGRTLLAAEKAMQER  
 BELSDCLRLAGSATTQEERAAMVAVKIGWEIDPRLVSPWLDLVAALRRGDLRS  
 GDAWVTKHLAQDRPFEAVESLTLAAGDAGEPGVEAELHFVRHVLHSCPLRAS  
 VPGSPNTPTRPRLRVRSPASVAGEVLKVLTEGGPDHTVAAEELRCRCRSETVV  
 EALRALALVYAGRPDRAAFWCDGLLQQAQTRSTGTSTVAILSGIRAEIALRQALPA  
 VEQYARSALSAILPGWGVAVSGPLAALVHSTTASGMSGVAGAWLQAVPDGMFTRH  
 GLLYTHARGHYHLAVRPAALDDFLTCCGLAKEGWMDVPTFLPWTSAQAQALN  
 PQARALVSEQLGPGSGTARAGVALRLILATGEPGRPALLESVNLLESQDOVE  
 LKALADRCQALDHVQAPAKMTARHARNVADACGGDALFRKLFRDEVPDRSGDPG  
 LTGEGQSASLTAAERVRVTAALAHGSHNRKIFITKSTVEQHLTRYVRKLVNR  
 ADLGDLLAGTNLGAAGTPAMAMASSAVAG"  
 10933..11271  
 /codon\_start=1  
 /transl\_table=11  
 /product="P450-like protein"  
 /protein\_id="BAB69191.1"  
 /db\_xref="GI:15823974"  
 /translation="MEVGGQTRAGEVGLCTLTSTANRDEGAFDTPDELRLRRDARSHL  
 SFGFGTHCIGQSARAEQIALETLLRLPGLRLAASFEELFRDGVGVGVKELPD  
 PCPALPDRP"

gene  
 CDS

11458..29898  
 /gene="olmAl"  
 11458..29898  
 /gene="olmAl"  
 /codon\_start=1  
 /transl\_table=11  
 /product="modular polyketide synthase"  
 /protein\_id="BAB69192.1"  
 /db\_xref="GI:15823975"

gene  
 CDS

11458..29898  
 /gene="olmAl"  
 11458..29898  
 /gene="olmAl"  
 /codon\_start=1  
 /transl\_table=11  
 /product="modular polyketide synthase"  
 /protein\_id="BAB69192.1"  
 /db\_xref="GI:15823975"







```
385..13257
/genes="myca ORF1"
/note="putative"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="BAA76543.1"
/db_xref="GI:4586928"
/translations="MSCRLPAGSGBEFMDLLTEGRDAVARQPDGSGWRAALABTGTFD
AGFGMSPROAAADPOORLRIELGWEALEHAGVVPATLAGSRTGVFVGVYADDDYATS
LMRTGTPDGYTVPGLNSMIANRLSYLLELRGFSFVWDSAQSSSVFVGHVLAACESLR
GESDAIVAGVSTSPSTPAWASGALSGDRCCTFDARANGYVRGEGGAANVLR
LSDALVDGRVHAVILGAVNDDGGGGLLAPRAAQEEVRLVDRSGVPGQAIVLR
ELRGTTGVPDGVFAAALGAGVADRRADDPLHGVSKSNIHLEGAAGITGLKXVL
CVWEGLPPTNLTQDTPGPIPLAQLRLVQTVTRWPDRADEGRITAAVSGFMGGTN
AHLVLAQAPRAVPAVPTTELPAVFWLSARSEALRQAEERLREHLDRDLRPL
DVGFSLATSDVPHRSVRPADGEPVAAAGSRLAEGRAVWVFPQCGAQTGMAAASLT
ESPVRARLITECAAAVTPFVDSLLDVLRAVPGSPRLDRDDVQVPAACFVWVSLAEIL
RSGWVPEAVVSGQGEIAAVAGAISELDGARIVALRSQLLTRERLHGLGMSVGL
PVAETERRLRWGRHLAVVNGFNATVAGDGDDELTECERDQARRVAVEYA
SHSPHYERIRDELMTALRPVTPQPSAVAFYSTADVGITDGERLDGAYRNLRQPVRF
ADLRQMSDGTGTSFVELSAHPMLVTLAEILAQETTAALCVTGSRLRDEGLRMLA
SAAEGWAGCLPDRWRKVFAGTGAERVELPTVAQORXYWAPDAERTAVPAAVDAPVR
PAASPEQATPAQTQQTIDILKTVLETVAIVQGLPSAEIVDPDLFRKQGLESTAA
VFNRLRARVGLSLSLPAULVDFPTTAVAGFLDQLHPRESQLVTVTPALVDDPEIA
IVMGRCRFPFGVSPEDLWVAVASVDVAFPPFDRGMDLESFDGDDPRAGTYCRE
GGFLDDAAGDAGDFGVSPEAVAMDPOQLLLEGGVVGAGGFSAGVGRSGRTGVY
AGVMOBYGPRLYESAEPHRYGLLTGTAASVSGRSVYVLGEGPAVTVDTACSSILV
AHLVLAQALSGEDNALAGVTVMTAPGI FVFSRQGLAVDGRCKAYSDSADGTCM
SEGUVGVVORLSDAVNAAGRVLA VRGSAVNDQASNGLITAPSGSQQRVIRAALAS
ARLSGVDDVVEHGTRTLDPIEAQALLATYGRGGGRSLILGSKNSIKVIAQAAG
VAGVWKMVAMERGVVPASLVRCAVEACFACQDSSGAVEVVEGACAMETGPRRAAVS
SFGVSTNAHVLIEQAPVPPVDASSGAPVASVADSSSVVPVLSRSEWGLRA
SAGRLAALGPIGLSGVSGVGVVDVGLSGRSVFGWRGVSLSGVGLGVGLGAV
VGSVGVSGRSVSGVGVFVPCGQOMVGMARGLCEGSEVSEAEVGVQGLVD
WPRVGTACREVGRLGRVDVQPLSFVWVALARLSRSGVVPVPGWVHGSQGEIAAA
VVSGLASVEGARGVGRSVI GARLGGAMVWVGASVGVVEGLGGLGGLGAVV
NPGQVVVAGDGGVGGVAGVGGVGGVGRARLLPVGSGHGLVDVVRDGLLVGLGV
RGSSTVFPFSSVGVGGVGGVGGVGGVGRNLRVFRGVVRLVGDGSSFSFEVSA
HPLVGAIEVLGEGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGV
VAGLGRVGLPTVPHDRYKLAPEPTAAGATAADGAAAGVGRVYRWGKALPV
TPGRLGRNVLVHGATGVGDQLSGVVDILTRHGATTEHTIDPAPSGVLAERLGL
AAAGEPVRVSLDLDLTLVQAMGRAGLDARIRAVTVGAVASDEVPSESGAAL
WAFGRVAALEPORMAGLIDLPVQADARALERLAGLTSSEENQAVRAGALXYTRII
RTAGARKWRPRVTLVGTGALGARVARLADHGAELVLVSRGAGATGAELV
AGLERTGVTVVACDVADRDGAGRAGGAGTAGGLPHRWARPDPGCPHARSGGG
VAPERLAATHLIDTLRDLDAFVLFSI VSGMAGQAAVAAANAALDALAEORRA
GLPATSANGLMAGMADGAGGCTFTRGVAMDPPDGTALALQALDSGTCVTVAD
VDWPSVWRTANPAARLFEIPEARQALADRAAEQSAERPSDLARLTLGITADRRL
LLDLICQAVALILRQDSGMTIPPSAFRASGFDLSLTALERNLNATGLTLPSTVY
FDHSPVALADFLDVLPRSGGAGVVEVAAAPDEPIAVIGVCYRPGVATPEQLW
DLILTERDAPLTPDRGMINDITFDPGCRVGRTYCREGFLDHAADFOAAPFGISF
REALMDQORLILLETSEALERAGIDPRISLRGSRCTGYTGMTHQVYARLHMEPEY
EGHLLTSGSAGRSVYVLGEGPAVTVDTACSSILVALHLAVQAILRAGECDLALA
GQVTVMATDCSMSGSPASGPRPVQGVLRQADGTGWSGAGVITLARLSDVDRGR
HILALIRGSTPTVGTGPRPTGRHSSSRLANAGLTPEPDAVEAHGVTGLR
GDTEAGALLATYQCRFPVQPVRLGTLKSNIGHTQAAAGVAGMKMVALRHGVLPA
TLHATPSHPDMNAGVALLTERTDWPDPNPRPRAAVSFGISGTNAHILLEOAPEP
TVPSERSAAGPLSVVPLSGRSEPALRAQARLDFLDGADVPDVGFSLATFR
SRLDRAVLLDGTATEALHALVSGDAGRAVLSGTPDTALFLFTGQSSQRLGMA
DGIRPEFPFAEALAEIAAMMDLQDPLLSVMAEPGSEATATILDRTEYAPALFV
EVALFRFESWGVRDVLGHSVGEIAAHVAGVMDLEHACRLVAARGLRMQMPADG
AMIAIRLPEFVRQRLDGLRHVSAVAVNGPESVVSAGHVSRSLADQLTAEGHRVR
RLNYSHPAFSVPHVDVLEEFKRAAAEVYGRPLVPVSVNTVNTVATEDLGAESYVVR
HLRSPVPHDGIRTLDRGATLVELGPDGVLGCLAAQCLDPPDEVALPALHRTSG
EVRSLNAAATVVRGHVDVNGFPETGRTGLSLAAQPLVYAFORQYRWASVSPGSDQTA
AVRFLNADHPLLGSLASVAGSDELITGRLSLAAQPLWIAQYRWASVSPGSDQTA
ASRAGAVGGRDRLDQTEAPLVVDRGAVQVQVLTGCDPDRGQVRVFRPPEGSTA
PWTRHATGNGDAAGSGRRPGVGVALPEGAVPLDVGDLVYRFAAQHLEHYDGVGTGR
ELWRADVEFAEVRLPDGAEVAAAFVHPALLDAALQPLAGDFVSRPADSVHLPPA
WQGMQVTSGAALRVIRIRAGESAASQVAVDGAFCVSLDALVLRPTTVRSRTTV
GPALLPLHLRWRREFGQATGTLTFLPTGAERYVGEATVTHQDILTALRAVQORGA
PPPRAVLAAGFSGPHRTAEAVRAVAAAAALDLVQRWLADDRHQDSRLVLVTERAVAADH
```

EDVPDLAAGAAVWGLVRSASEHPGFALIDHDTQESRALLPAALGCAQLARQGT  
VRVPAALAPPTESGADTRELAPDGTIVTCAAGALGALVVRHLVABHGVRLLLLSRH  
GREAGMABLAEDLAQRGVTVVIAADVADRTGLAAALDAVPDEHPILTAVVHARGVLD  
DAMVTCGLTADRLDEVLVRPKVDAALHLHLELTADRKLAALFVLSAAAGLLGNPGQASVAA  
ANTVLDALAHBRANGLPATSLAWGLGGPNGMAANLAAHLRLARLYGTAPLDGDEA  
RSLLDRLALTRDRALLVPLRLTRVDRQADPADVEILRLPLVVFVADGGGSPATGNS  
LEAFRAQALAEDDGTREHLLDVRQVAVRIGHRADRVVDPKRPFDJGFGDSLLAVE  
LNRLOQATGLPLPATIGFPHPTPTQVPHLAESLISRPEDGPEQALTLGLDAVARAL  
TKLGDDIRRVVVRRLAEALLTAGAEPLDANGGTTGAGVDDGHDDTALVERLSGAS  
DEDLAPIDBEQI"

ORIGIN

Query Match 5.2%; Score 26; DB 1; Length 17101;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 GTCCAGCGCGTGGAGCGACGCCAC 410  
Db 8656 GTCCAGCGCGTGGAGCGACGCCAC 8681

RESULT 36  
AF016585 41097 bp DNA linear BCT 07-DEC-1997  
LOCUS Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)  
DEFINITION gene, partial cds; polyketide synthase modules 1 through 7 (nida)  
genes, complete cds; and N-methyltransferase homolog gene, partial  
cds.  
ACCESSION AF016585  
VERSION AF016585.1 GI:2558836  
KEYWORDS Streptomyces caelestis  
SOURCE Streptomyces caelestis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 41097)  
AUTHORS Kakavas,S.J.; Katz,J. and Stassi,D.  
TITLE Identification and characterization of the niddamycin polyketide  
synthase genes from Streptomyces caelestis  
J. Bacteriol. 179 (23), 7515-7522 (1997)  
JOURNAL 98053867  
MEDLINE 9393718  
PubMed 9393718  
REFERENCE 2 (bases 1 to 41097)  
AUTHORS Kakavas,S. and Stassi,D.  
TITLE Direct Submission  
Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park  
Road, Abbott Park, IL 60064, USA  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1. 41097  
/organism="Streptomyces caelestis"  
/mol\_type="genomic DNA"  
/strain="NRRL-2821"  
/db\_xref="taxon:36816"  
complement(<1..396)  
/gene="nidi"  
complement(<1..396)  
/gene="nidi"  
/note="ORF6; similar to tyli"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450 hydroxylase homolog"  
/protein\_id="AAC46023.1"  
/db\_xref="GI:2558837"  
/translation="MVDSVTGPMELSKDANAKELLDFSHNRTHPHVFWDEGRQAVQV  
FPRYDDYLTVNHPHFSSDFTEVAPTEPELEMILPGTIGALDPPAHGPMKLVQAF  
TPRRMAGQQRIRVIAELLDRVRGQKTA"  
627..13649  
/gene="nida1"  
627..13649  
/gene="nida1"  
/note="orf1"  
/codon\_start=1  
/transl\_table=11

gene  
CDS  
gene  
CDS



```

/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translation="MAGHGDDATAQAQAQAEKSDGSDAIAVIGMSRPPGAPGTAEFW
QLSSGADAVVTAADGRRGTTDAPDFDAFGMSPREAATDPOQRLLVLELGEAL
EDRGIPVSGEASVFGAMNDYATLLHRAAGTPTTATGLOHSMIANRUSYFL
GLGPIVVDTGSSSIVAVALESRGTSIGLAGGVNLAEGSSAEMRVSGL
SPDGRCHTFEDARANGVEGGAIVLKPADALADGDRVYCVVRGATGNDGGPGL
TVPRAGCEVILRAACDQAGVPRADRVFELHGTGTAGDVEAALGAVYGTGRPAN
EPLLVGSVKNHLEAGAGIAGVAFKALCHERALPASLNFETPNPAEPLRUKV
QTAHAALQPGTGGGELLAGVAFMGCTNCHVLEETPGGQPAETQOACLFASP
MLLSARSEQALRAELRHLEDSCADPLDIAYSLATTRFEHRAAVPCGPDRL
SSALAALAAQTPRGVIGSTADGRLALLTGQAHPGMGOELVYTTDPHFAAALDE
VCEELQRCGTONLEWMTDPODPLDRTVTPQALPALQYLTTLTARGTQOHLVL
GHSVGEITAAHAGVLDLPDAARLITABAHVMGOLPHGGMLSVQAHEHDLDAHCH
GVEIAAVNGPTHCVLSGPTALEETAQHLRQCNVRLHVKVSHAFSHSALMDPLGAFR
DTLNLNYPPTPLISNLQGIADPNHLCTPDYNIWHARHTVPADAVQTAHQGTT
TYLEIGTPHTLTLLHHTLDNPTTPTLHREPRPEPTLTOAIAVGVRTDIDWALVC
GASRPRVELPTYAFQRTHWAPGLTPNHAADRPAAQPORAMAVGVSREALVRLVG
ETLADVLGDPDEVALDRPTSSGLDSMTAVELAGLLGTAGVALDPTLYELPTPR
AVADHAKTLLGESADADQVNGRTGEAEKAGDPIAVIGICRPGCGVATDPLWE
LVASGDTAISTPFDGMDLGLDPTDESTPGKSTVRGGFHDRAQDAEFGLSPR
EATAMDPOQRLLSTSWELBRAGVPSLRGGRTGVFVTTAYEPGRHLEGTDYE
GFLTGTGTVASGRIAYALGTRGALVTDTACSSILVALHVAQSLRGGCDLALAG
GTTVMGPGMVFESRQGLAPDGRCKAFSADADTAWAEGVMLLVERLSDAERLGH
RVLAVRGTANQDGSNGLTAPSGPAQOQVIRDAISDAGLSADDIDAVEAHGRTAL
GDTPEAGALLATYGHKQTPVWLSKSNIGHTQAAAGIAGIKMWQALRHDTLPTT
LHADPSSKQVMDAGLQLLTDARWPADPRPRKAGISAFGVSCTNAHVLEBPPAP
BEAPATQPPPTVAMPLSNHTPTALRAQARLDSWLHHTQADPODIAHALATRTFH
KHRAVITGRTEHLKHTLDAIOGTAPHPRLTLFTGQCAQHRGMGOELYADPH
FAAALDECEELQRCGTONLEWMTDPODPLDRTVTPQALPALQYLTTLTARG
TOAHLVHSGVEITAAHAGVLDLPDAARLITABAHVMGOLPHGGMLSVQAHEHDL
DQALANTHGVIEIAVNGPTHCVLSGPTALEETAQHLRQCNVRLHVKVSHAFSHSALMD
PMLGAFRDTLNLNYPPTPLISNLQGIADPNHLCTPDYNIWHARHTVPADAVQTA
HTDRTTYTQHLDTLTPPFQDRYFMEFVRVAQVSGPGADRLRYRVWMAAPREDT
VPSAREWLLGSGSDPADLIGVVEOLVGHGAQVRLDIPAPACRALLPCELSPEVR
QSHNGEQLGSLRSPRVAEVALIOALGDAGADPLMTATGVAQAQGAEPVGGGA
QWMLGQVAGLEADRWGLVDLPADPAALRGRLARVLVSNAPNOVAIRASGVFVR
RVVPAPGRVNDWAPSGTVLITGGTGLGSQVARRLALAGAPHLILLAGRRNSGAAE
LVDELATLGAETVAACDAADRDLVSLIATPEHRPLTAVLHAAAGVLDGLDGLTP
ERIDAVRAKATAAARHDLTADLIDAVLPSSIVGVNGGQATYAAANAALDALLA
HGRARGORATSIAMGWAGSMAAGDGAKSALRDIIVLREPEALQALAAVARDDET
TLVADWNEFELSSASRRFOLPLFGRIPARAFAVAGSADGSDPDPVWLSPOLSAA
ERQALLDLVIRTEASVRLRHGSATVDAETAFRACGPSLTILVELNRILASATGLRS
NTLLDHPPTPAAYSLGELGTEPSAASHGADAGRTVVASDPAVWAGACRIFG
GYATPDDLWELVAGSDTAISTPFDGMDLGLDPTDPTGKSTVRGGFHDRAQF
DAEFPGISPREATAMDPOQRLLSTSWELBRAGIAPHTLRTTRGVFTGASQOBYG
OSREAADKYGGHLLTGLTASVMSGRVAYTLGQGPALVDTACSSILVALHVAQSLR
EGCEDLALAGSSTVMSPTTVFVSQRGLAPDGRCKPADTADGTAWAEGVGLLVE
RLSDAERLGHVILAVRGTANQDGSNGLTAPSGPAQOQVIRDAISDAGLSADDIDA
VEAHGTGTPGDIPEAEALLATYGHPRQTPVWLSKSNIGHTQAAAGIAGIKMWQ
AMHGTLPRLTHVDAPTQVDWEAGSLQLLAEARWPADPRPRRAGISAFVSGTNA
HVLEBPEAPEAPATQAPPVIAWPLSAHTPTALRAQARLDSWLHHTQADPODIA
HALATCTTHFKRAVITGRTEHLKHTLDAIOGTAPHPRLTLFTGQCAQHPGM
GOELVYTPHFAALDEICEELQRCGTONLEWMTDPODPLDRTVTPQALPALQY
ALYRTLARGTQOHLVHSGVEITAAHAGVLDLPDAARLITABAHVMGOLPHDGM
LSVQAHEHDLQALHGHVIEIAVNGPTHCVLSGPTALEETAQHLRQCNVRLHVKV
SHAFSALMDPLGAFRDTLNLNYPPTPLISNLQGIADPNHLCTPDYNIWHARH
TVRFADAVQTAHQGTTTYLEIGTPHTLTLLHHTLDNPTTPTLHREPRPEPTLTA
LATLHTGHTTTPHPSHIAQORVSLPAYFORAYVWMSAAHIGRSDAEPLATGLA
RODHPLETCATSVISGATLLGRVSLAAHPWADHTVAGAVLPCTAPADLILHVAE
FTGSGTVBELTLHVLVLAQAVQLVLEAPDERGRAATVAARPEETGAETHH
AAGVLASGAPAPADVMDAAGAMPFGAEAVDVDELAAEFAAGYDYGAPTGLTG
AVWRGEFADVLRLPEAVAGSBEERFGVHPALFADALHPWRVCGPWSQESDADGG
QSTLLPFSMQGTLTYGTGAQVRLSLPDSGAFSVAADAPHRVLAFLPADLHVLRRVP
STTCGALPAAPMWYHIDQVPDSTSLVDTTRYAVAGVADADETARVLRPLDADHTAHA
DLAALBAADVAVTPFGPALVVALIPGSPPEAPATSAQDPVRATLHGLSLVQALIE
SBRLLTGTTLAVVTRGAVETQDVPDLAAGALWGLVRVAQSPGPTGRFLDIVDSDPE
DRAALATATALASDQIALRAGVMLPALTPAAAGASAPADAAPATGMGMDVTLTGC
TALGRVAVGLHARRHGRVRLMIVSGNPAESALRELTELTVHGTTSFVSCDLAP
ADVRAKAVGVPDHPHTLVVHTAGVLDGALARTPTERTIDTVLRKAAVRLHLEATL
GHPLRAFVLSAAGLGRGQASVAAAANALDAPARLRAESGLPAVSLAWGLWDBRE

```

```

GMAGGLDATALRRRLRRDGLLMPMPADQALQLDQALTTTYRDGPALLVPLLLDSTARLT
AGENGPAVPLLRSLVPSHPRRHTTTATSVRDESGPSPGARSALAAALPAKERMVTLID
TVSQVAGVLGHAASAEIDPSPHEIGFDSLATLELRNRLGRVLGRLTFLAPDHP
TPRKVAWINNELPAANPALGTSAEDPVLGIDOMARALALMDAEDGRADVORQLVG
LLTALDPAAGADSTRAAATTVARLDBATDDEIFAFLEQLEQEPDPLGAPLDERL
DEQ"
13693..19212
/gene="nidA2"
13693..19212
/gene="nidA2"
/feature="orf2"
/codon_start=1
/transl_table=11
/product="polyketide synthase module 3"
/protein_id="AAC46025.1"
/db_xref="GI:2558839"
/translation="MTAENDKIRSYLKATATLHRTKERLAELSAERHPEVAVVCMAC
RYPGVATPDDLWELVAGSDTAISTPFDGMDLGLDPTDPTGKSYVRHGGFLHD
AAQDAFFGIGSPREATAMDPOQRLLSTSWELBRAGIAPHTLRTTRGVFTGMYD
DYGRCCFPLPEVGVVNGSAGSVASGRVAYTLGQGPALTVDTACSSILVALHVA
QSLRGGCDLALAGSSTVMSPTTVFVSQRGLAPDGRCKSFADTADTAWAEGVGV
LLVERLSDAERLGHVILAVRGTANQDGSNGLTAPSGPAQOQVIRDAISDAGLSAD
DIDAVEAHGTGTPGLDPIEAGALLATYGHPRQTPVWLSKSNIGHTQAAAGIAGI
KMWQALRHDTLPTLHREPRPEPTLTOAIAVGVRTDIDWALVC
GTNAHAIIEEPTTLIEAPVDETPVPAHTPPIVAMPLSAHTPTALRAQARLDSWLH
TOADPODIAHALATRTTHFKRAVITGRTEHLKHTLDAIOGTAPHPRLTLFT
GQGAHPGMGOELVYTTDPHFAALDEICEELQRCGTONLEWMTDPODPLDRTVET
QAPALQALYRTLITARGTQOHLVHSGVEITAAHAGVLDLPDAARLITABAHVM
GOLPHGGMLSVQAHEHDLQALHGHVIEIAVNGPTHCVLSGPTALEETAQHLREQ
QMLHGVIEIAVNGPTHCVLSGPTALEETAQHLRQCNVRLHVKVSHAFSHSALMD
NYRWHLKVSHPAFSALMDPLGAFRDTLNLNYPPTPLISNLQGIADPNHLCTP
DYLHARHTVPADAVQTAHQGTTTYLEIGTPHTLTLLHHTLDNPTTPTLHREH
PEPTLTALATLHTGHTTTPHPSHIAQORVSLPAYFORAYVWMSAAHIGRSDAE
AAGYTAGVGHPLLAATVEPTDPTVMTGMSARALPDLHVAWDRVAGVAPTALEIV
LEIGERICGLPCIGELVLEAPLSKGDADAVVTVVGLSPECTREVSLSRPEDEAVM
TRHAGRLARAGSEMGSGVFAELSGVPPGQAEVDISGQVELFAAGVGGGAFRG
LRAWRRELDIEFAEVLDPDDHATDAGYGLHALLDAAALHPITALDPLDGGHGLLPF
VMTDVRPENGHTLRLVRTRAQDPADGQDRPGGVSLAAADRGVDFVFAVSLVL
RVSATOLPAAPFVPLHRLDWPIPGFTSTGAGSDPSVLDLITELTADHTVREVL
EAVOAHLEPEVPGDGRVVLTRGAGPAGSDADAPAGAAMGLVRAQSHPGKGFVLI
VGTDESSRALPAMVAVGDVDPQAVVRQGRPSVPLVPLTADTEAPEGPGSGAGL
LDGPTDVLITGGTGAALAVTARHLVARHARHLILVSRGTDPQAAELASLGAALGA
DVEFFRACVTDRAAVORILLAGIPAEPLSLSCVHTAGVLDGDIISACTPERRAAVILGPK
ADGHALDELTRDRERSVPLVLFSSVSGVLTGSHAGYAAANAFDALARARADGHP
ALSLCWMGTGTGMAAGLDADATARIIRSGIAPLDAETAMOLFDRALARPEALLPVR
LDHALARAAAGATLPEVLRLDLAGVPSRAHPAGSEGTATSTGTGSLAERLARLTV
ERLPLLELVRTEVAALVGHGDPGAGVABERSEFKDAGFDSLTAVDLNRNLNARTGLP
ATLVSDHPPTPLALAEALLAGLDEVEADGPEDLAAAFDRLAGLALDADRAARYGAE
RLRSLTALTADPATGTGAQAGPAGDGAITVTVTRKRSASDDELDFLDYDSEFR"
19269..30233
/gene="nidA3"
19269..30233
/gene="nidA3"
/feature="orf3"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 4 and 5"
/protein_id="AAC46026.1"
/db_xref="GI:2558840"
/translation="MSATNEEKRLRYLRAMGDLHSARERLAELSAERHPEVAVVCM
CRYPGVATPDDLWELVAGSDTAISTPFDGMDLGLDPTDPTGKSYVRHGGFLH
DAAQDAFFGIGSPREATAMDPOQRLLSTSWELBRAGIAPHTLRTTRGVFTGMYD
HDYGSQVAGSADAGSGLGPGASVGRVAYTLGQGPALTVDTACSSILVALHVA
VQSLRGGCDLALAGSSTVMSPTTVFVSQRGLAPDGRCKFPADTADTAWAEGVGV
MLVERLSDAERLGHVILAVRGTANQDGSNGLTAPSGPAQOQVIRDAISDAGLSAD
DIDAVEAHGTGTPGLDPIEAGALLATYGHPRQTPVWLSKSNIGHTQAAAGIAGI
KMWQALRHDTLPTLHREPRPEPTLTOAIAVGVRTDIDWALVC
GTNAHAIIEEPTTLIEAPVDETPVPAHTPPIVAMPLSAHTPTALRAQARLDSWLH
SIADSGALGHGLAVGRALPHRAVLGDCGAPDLAALASGESVDVVTGSAADVGR
VAFVPPGQCAWAGMGAELLDSSPVFAEALCEALPEFVDMSLTDVLRGAPAGVL
DVVTVQVPTFAVVVALAAMWRMLGVEPAAVVGHSQGETAAAHVAGVLSLEDAARVA
LRSQILATLRELGRGSAVLAADAVESAGAEAGGVDRVDEIAVNGPTHCVLSG
PGASLGLVGLGEGVRRIDVYASHSRHVEGIDELAAVLAGRLPAGRPAGVYST
VEAEPDGLDGTALDAGYWRNLRQVRVRESALRAMLADGVDAFVECSPPHVLTVPRQTL

```

EDAGAGAVAVGSLRRDDGGLRRFLTSAEAQVAGVPVDWAAALCPRAGWDLPTAYAFQ  
ERYWVAPEAPGPAAGSAATGPAATAARSQSGEDAASAYGDRLAYHYVAWQGLLP  
CADAWRPGLRLIIVPAGEREAADVSPQAIASFQGTGVRVYVADPLSTGRDGLRALIA  
PVVAGEEPTGAVSLGLCTGDHPGHPGVPGTGTLALVQALADGAVAVPLVWTRG

Query Match 5.2%; Score 26; DB 1; Length 41097;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 GACGGCGCTGCAAGGGCTTCGCCGA 167  
|||  
Db 24315 GACGGCGCTGCAAGGGCTTCGCCGA 243

RESULT_37	
AF263912	
LOCUS	123580 bp DNA linear BCT 24-MAY-2000
DEFINITION	Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster, complete sequence.
ACCESSION	AF263912
VERSION	AF263912.1 GI:8050835

SOURCE	ORGANISM
Streptomyces noursei	
Streptomyces noursei	
Streptomyces noursei	
Bacteria; Actinobacteria;	Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomy-	cetaceae; Streptomyces.

AUTHORS	Bräutast, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Ström, A.R., Vallia, S. and Zotchev, S.B.
TITLE	Biosynthesis of the polyene antifungal antibiotic nystatin in <i>Streptomyces noursei</i> ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway
JOURNAL	Chem. Biol. 7 (6), 395-403 (2000)

2007/06/11  
 2 (bases 1 to 123580)  
 Bratrasæt, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R.,  
 Valla, S. and Zotchev, S.B.  
 Direct Submission  
 Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim  
 N-7489, Norway

```

FEATURES
    source
        Location/Qualifiers
            1..123580
                /organism="Streptomyces noursei"
                /mol_type="genomic DNA"
                /strain="ATCC 11455"
                /db_xref="ATCC:11455"
                /db_xref="taxon:1971"
                complement(46..783)
                /gene="nysF"
                complement(46..783)
                /gene="nysF"
                /function="presumably post-translationally modifies ACP domains on PKS"
                /note="putative 4'-phosphopantetheine transferase"
                /codon_start=1
                /transl_table=11
                /product="NysF"

```

gene complement (867..2684)  
PAPTPTDRREHRLTHLVNGLPRLPTFG"  
LTQRELFDSAEIVIDPTQGAFTARILVPGFLIGGRRVTVFPGRWHSHTPALLTAVVHL  
AVSIDAEPNGPLPAGVYNATALPSEPHILVALAAHRDPVVDRLLLFSKESVFKAMPT  
FTTVRLARRALRGHPORAILPNRRGAQWPGPVGSMTHCAGYRAAVSAFELSA  
/translation="MIELIIPATVATVAAAYDDPRCDRLLSEREFVIARAVSREORE  
/db\_xref="GI:8050836"

```
Complement (60 / .2684)
/ gene="nysg"
/ function="presumably involved in efflux of nystatin"
/ notes="putative transporter (ABC family)"
/ codon_start=1
/ trans1_table=11
```

gene	CDS
<p>/product="NysA"</p> <p>/protein_id="AAF71763.1"</p> <p>/db_xref="GI:805083"</p> <p>/translations="MASPDLEERTAPREARELVGLRRPHRSVALAYSMVGGLV NARGPLLLGRVTLIADIDVGLGVGPAPGIDFAAIGRLLLVLLALVVASLFWMAQGR LVASAVWRTIHELRDAREKRLRLRHRDFRQAGELLRLRTNDINLQOTLQOTLAE LITSIFSLTLMVLMVLSPLAVMLLSVPVSALIAARIKRAQPHYAAQWNSANGVL NAHVECTGHALIKGFDRAAEERFDACNDVAYRAAKAQAFASGAMSPVMWVFANL GYLVAIVAGAKWNGTFLTGDDVQAFILYARQFSQPIVETASVAGRLQSGIAFAQV TLIDPAQAPDPLRQPTPARAECGRVEFDVSYFRYSPDPLIENLSLTVBPGSTVAIVG PTGAKTTLGNLLMRFFPSPDPAEGRVLTATMTRDRLRSRGLVIQDTWLPFGGIIA ENTAYGAPGACRADIEBAARATCADRFIRTLPGQYDTVLDDSGTYSAGEKQLLTVAR AFIARPAVLVLCATSDVSTRTREVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIIVM DAGRIVBOQTHDOLLCAQGLYAKLHAARTHTPTAGAAG"</p> <p>/gene="nysH"</p> <p>complement(2662. .4416)</p> <p>/gene="nysH"</p> <p>/functions="presumably involved in efflux of nystatin"</p> <p>/note="putative transporter (ABC family)"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="NysH"</p> <p>/protein_id="AAF71764.1"</p> <p>/db_xref="GI:805083"</p> <p>/translations="MLRLRLAQLRPYAWATAALVQLVQILGTLLPTLGAALIDQ GVVGGDGRITELGVGVGVVALVQIIAALGAALAAARTATAMGRDILRSALFREILDF AREIGRGTSLSLTRSVNDVQVQNLQAOTGFGIWCAPMLCGSLVLLARQDVPLALL LVALIVVAVCFGILLARMGTLVARMQTLIDRLGRLLREALTGVRVVRVFRVDDHERA RFAQNTDFAVRRVRRGLRLATMLPVLLMNGFTVALLWTGSHRVDAGRMPTGSLSA LLSYLSLISVVMVLAFFVLSVPRVACRIAEVLDTGSSVAPPAPQFVRPGAGRI ELCAAGVRYPCAEPEVRDVLTVPEGERFVILGTSQGTLLNLVRLADATGAV RVGCTDVRELTAAATAAGVGFQVQPVLYESGTVASNLRGEPDATABEELWEALRYAQ ADFARMNDELTAELITCGGNSVGGQRQLSLARALLRREPIYLFDDCSALDQATDA ALRTALPYTAGATVITVQGSAGRADRIIVVLDGRVVAQVQVLTLLTSPTYREI ALSQLEEEAAHGLACRP"</p> <p>4714. .5748</p> <p>/gene="nysDIII"</p> <p>4714. .5748</p> <p>/gene="nysDIII"</p> <p>/functions="presumably involved in mycosamine biosynthesis"</p> <p>/note="putative dGDP-mannose-4,6-dehydratase"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="NysDIII"</p> <p>/protein_id="AAF71765.1"</p> <p>/db_xref="GI:805083"</p> <p>/translations="MSKRALITGITQDGSYLAHELLSQYQVWGLIRGOANPKRSRV SRASELDFIDGLMDQGLSVASVDTVQDVVNLGAI9FVPMWQOAEILVTEVMGQ VRLMLEARHIVYVNSGLSTRYSRGOIRFYQASSEMFGKAAETPQRETLTLEHRSYVG AAKAYGHVITPNYRESGFMVYSGMLFNHESPRGQEFVTRKISLAVAILKEGLQDLKL ALGNLDVRDWGAGYGVYRAMHLLQDDAGDDIYIGTQGMHRSVRDRAVITAFBEVGLNW EDYVIDFDLVRPAEVEVLCAWSAKAQDLRGKPKDVPDFTLMRMVMVDSLIAQVSRNQ YGDVLLAANW"</p> <p>5930. .34363</p> <p>/gene="nysI"</p> <p>5930. .34363</p> <p>/gene="nysI"</p> <p>/function="responsible for condensation steps 9 to 14 in the nystatin polyketide backbone synthesis"</p> <p>/note="polyketide synthase"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="NysI"</p> <p>/protein_id="AAF71766.1"</p> <p>/db_xref="GI:805084"</p>	<p>/translations="MDNEOKRLDYLLATADLRTRRRYVHLESAAQEPVAILGMTCR YPGVRSPELWRWEAGEHGVTPPTDRGWDLEALAAAPTASGGFLHAPDAPDFDFF GISPRFVAMDPQORVLESMAEFERAGIDPVSVKSGRTGVIFGHMAQDYRVGPADP AGEFQITGNTGYSLGRISYTFQTVGFAVTVPTACSSSLVAVHLATQALRAGECTLAL AGGVITMSGPTIEMRQGLISAGDCRSFGDTADGTGWAEGVLTILVRLERSDAVN GHEILAVRGTAVNQDGSNAPETAPNGPQQQVIOQALVNARLACGDIIVWEAHGHTI TIDGVEAALATAYGONAPETLGVSKNSLSEHQAACGAGVGIIDVWAMVHGCTI</p>
gene	CDS

PRTLLHABEPTTHVDSQAVRLRLTDTTDPATGAPRRAAVSSFGISGTNAHTIIEQAP  
 EPOEDATAOQADGTPATAPVVPVVLGGTPTDARLQQAALRAALDTCRRP  
 DLLLALSLATRAAGFHRVALLATDHPALDGLTALADADDPAAPAWITGTTRAET  
 RLAVLFTQGAQRAGRELAAEPAPAFATDAALDAETPHLDRPELVGTDAALI  
 DRITAGPALFAVEVALYRLTESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA  
 AGRMLQALPGGAMIAVOASEADVAPLLAGHEDQVAIAAVNGSPVAVLSGAEATVTA  
 LAEQLAADRGRRLIRVSHAFHSPLMEPLMDAFRAVEDTLQPLLPVSNLTKGA  
 TAOULTADYVWVRHVRADGIDMLARHDTTAFLELGPGLSAMAQDCLDAADA  
 DAVTLPALACRPREHTLTALAGLHVHATLDTWCFCAGTCAARRDTPVFAFORRY  
 WPKALQSTADLRSVGLAAHPHLLSAVSLADAGTLLTGLSRQTHPMLADHVRG  
 TTNLPGTFLAVAGVEGCDRVBEELTLAPELLPQGGVQVQLMHTNPDVSGRR  
 VVHARPDGDDTTHATGVTITADASROLPASSEGGTGLAGDHPALDAQWPP  
 AGAEPFLDGLDGLGFGVFPVQGLRAARNGGDVYAEVELPERGSDAEAFGL  
 HPALLDAALHAAPFTGLGERGGGLFSGWEGVSHAGGATLRLVLTAVDADALVY  
 ADGTGAPVLSVSVATVATQOLDTAARVADALFLDWTVPQVATDPGPVALLGA  
 DPGGLLTHAGPADAPDLAALAAAGVPTTVVLISLACTDADPARSAHRCAAE  
 ALAAVQTLMDHHERFAARLVFTKATVGRDVAIAAAVGLVRSQNSQSPFCALVDL  
 DPGAVGAALVAALVSGEPQLAVRGDVLVRLVRLPTEVAGADGTGGVGGSG  
 VSFSGEGLVLTGGTGLVGLARHVAEYGVDRLLVSRGSERAVAGGELVAELAGY  
 GARVRVACDTRDAAVVELVGGHVAHVAAAGVLDGMVGLTGERLSAVLRKVD  
 AVHLHEATRGDLDADFVVSGLAGVSGPGQAAVAAANAFDMLTERRRAEGLPKLS  
 LAMGPEQSGGWTGLTDVAERLARSGVPLSVAQGLALPDAAVAGTATCTVPRLD  
 LPVLARGSEVPLLSLRRLARRAAVAGSATAGNLAQRLRLDEGDRMDLVRG  
 QVALVGHATGVDAGRAFDGLGDSLTVAELNRMLNTVGLRPLATLDFDIPVRH  
 LATVYDLRLGDAEVATVPAAVADDPIVIGMACRYPGVSSPEDLRLVLTGRT  
 DAVSGFTNRMGDSVSLHDPDPHPTGTSYTRSGFLHEAGFDPGFFGMSREALATD  
 SQORLLSSWEALIRAGIDPVSLRSGRTGVAGVMDYSAMLASPEFPGQSGSS  
 PSLARGVAVTGLRGPATVDTACSILVAMWAOALRSGECGLALAGGVVMTSTP  
 AVVDFAORGLSPGRCKAFADADGVWSEGVGVILRQSDAVRNGHEILAVRG  
 SAVNGDASGNLTPNGSPQVTRQALASGLTADGVVVEAHGTGTGLTGPIBAQA  
 LLATYGRDRPERPILGSLKSNLGTQAAAGVAVIKVLMHRHGVVPTLHVDAQ  
 SHVWSEGAVELLSQAAWPTGVRVRAVSGFSGTGNVHVIVPEQAPCAKAAAGA  
 ARTPGAVPVLISGRGALRGQALRQALPDALVDVALSLATRSFEQRAA  
 VVAORDOLIASLGLAADRDPVAVEGEAGRTAVLFTQSGORAAAGRELHEVO  
 SHRTDQVAVLSSQAAWPTGVRVRAVSGFSGTGNVHVIVPEQAPCAKAAAGA  
 PEFAAAFRAVCAVPLDRLRREVRFAEDGSDAALLDETGTQPALFAVEVALFRL  
 VESWGVDFVAGHSIGETAAHVAAGVLTLEDAKRLVARATLMQALTPGGHAIQAI  
 TEDETAHLDHDTVAIAVNGPQSVISGDEEAETIAATFAERGRKRLRVRSHAFHS  
 PRMDGMLDAFRITVAEGLVTRAPILVSLDTRADDAAEVCTAAYVVRHREVRFPAD  
 CVRTLRDAGATFLELSDGDLTMAEDTGLDHDDELVPLMRAGRAELAAATALAR  
 LQVRGDVDMWALAGTGARRDLPYAFQHAYVMPQLPTPAALAAADPADQALWAA  
 VERGDARELILGEGQDLPDLSILLPALTSSWRGNQKHLDLTLRIVKVEVTRLSKP  
 TAPLDVGTWLLVASDADATDALLDGLADALSHGARVRLLLDDSDCADRAVLATRL  
 ARTADVDAATVLSLPLDERDADDCPLTGLTALTVLQALADTGAQRLMTATRG  
 AVSTNPADVTHPVOAAAGLGRVHALBHPRLGVLDPQVDFERAGORLAGILAVK  
 DAPDGEDVALRATVSGRRLVTRVLEALPTAAEFTATGTVLTGTGGLGAELVARWL  
 ABAGAQLVLTSERPDPAAGAELELGEYGPSVSVACDVADRDALAAVLTALPEE  
 LPLTVHVTAGVGHVGLDLSLTAEPAGLTAALAGAHLDALLADRELDLFFVLPGSI  
 AGVWGSQSAAYGAANAYLDLALHRRRAGLAATSVAMGPWAEAGMAADAVSETLIR  
 QGLGLDPAAPANTELRVAVRQDVTVDVDMQRYAPLTSARPSALTAGLPEVRAL  
 AADERTQDAGTESEVTVVRALAEPEQLRLTDLVLTESATVLSHSSADAVPEGRAL  
 RDVGFSLTAVELRLKLGATGLSLPSTWVDPYPTLELAQVLRABILCAVLEVAGPV  
 ATGGADDEPIALIGMACRPGVSVSPQOLWLVASGTDAISEFPVNRGQWGLHPD  
 PDRGTYTQGGFLHDEDFPTFGISPREALVMDPOORLLLETWESPERAGIRP  
 ETIRSLTTLFTGSSYQYEGLAGDGTGHTVMTGSSPSVLSRSLSVFLEGPATVD  
 TACSSLTHVHLACSLRNGENSLAVAGGATMTTNPFIAPRORALAKDGRCKAFS  
 DADGWTGLAEGVVLRLDAQNRHPLVAVLRGSLNQGASGLTAPNGPSQOR  
 VTRQALANARLAPGDIALEAGTGTPGLDTEAQALETATYGRDRDPESALILSVKS  
 NIGHTQSAAGIASIVIKWVALRHELPPTLHADAPSSHVDWNSAGTVRLTQARWDET  
 GPRRAASFGISGTNAHTIIEQAPVAPVPAEVRPAVPTPAAGVFWVPTVARSAA  
 ALURGERLALHAEITGTLPAAGPLDILGLSVARSARFHRVAVVPPPTPLAALR  
 AVATDGPSPVAVGADVAGRTVFPFGQSGQWNGSQQLDESVAFAERIAECRAAL  
 ABEFTDMSLVLRGVGAPSLERVDVVPASFAVMVSLAALWRSQVLAFAVSHQSG  
 ETAAVVSVALSRIDGARVVALRSQAIGRALRAGRGWMSVALSVDPRLVEFEGRV  
 SVAVNGPSRVSVFPGPEALDALHRLTADITARELVAVDYASHVQVEDHELELV  
 LAELAPRTSVFPTGMDLTARDMDITFRNLRGRVFRFADAVDLALHARYAFV  
 EYVSHVPLSMAQVQEAAGVAPVAGTLTRDQGGTDFLLSAAVFGVGVDDWAGL  
 FEETGASRLDLPYAFQHEHWAVPAPPAEVAADDDAAFTWAVGSDVSLATLALG  
 TDDSVAAVLPALTSWRARRDRSTVDVARYVWAKPLGGLTPHPSLITWLLVTDAG  
 IODSDVALETVAGVRLVDEECVDRVLEPERLAGAEDVTGIVSVLAAARBTADV  
 PGTSVLGTALVTALLQALGDABIDAPVALITGAVSTGAEDELTAPOVQVGTIGWT

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 142 GACGGCCGCTGCAAGGCGCTTCGCCGA 167  
 Db 51497 GACGGCCGCTGCAAGGCGCTTCGCCGA 51522  
 RESULT 38  
 LOCUS AX2111739 125401 bp DNA linear PAT 06-SEP-2001  
 DEFINITION Sequence 35 from Patent WO0159126.  
 ACCESSION AX2111739  
 VERSION AX2111739.1 GI:15523950  
 KEYWORDS  
 SOURCE Streptomyces noursei  
 ORGANISM Streptomyces noursei  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1  
 AUTHORS Zlotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,  
 Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and  
 Gulliksen, O.M.  
 TITLE Gene cluster encoding a nystatin polyketide synthase and its  
 manipulation and utility  
 JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;  
 Norges Teknisk Naturvitenskapelige Universitet (NO); STIFTELSEN  
 IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);  
 ALPHARMA AS (NO); SINVENT AS (NO); Zotchev, Sergey Borisovich  
 (NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);  
 Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein  
 (NO)  
 FEATURES  
 Location/Qualifiers  
 1..125401  
 /organism="Streptomyces noursei"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1971"  
 /note="APCC 11455"  
 ORIGIN  
 Query Match 5.2%; Score 26; DB 6; Length 125401;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 142 GACGGCCGCTGCAAGGCGCTTCGCCGA 167  
 Db 51905 GACGGCCGCTGCAAGGCGCTTCGCCGA 51930  
 RESULT 39  
 LOCUS AF521085/c 132544 bp DNA linear BCT 23-MAY-2003  
 DEFINITION Streptomyces nanchangensis NS3226 polyether ionophore nanchangmycin  
 biosynthetic gene cluster, complete sequence.  
 ACCESSION AF521085  
 VERSION AF521085.1 GI:31044135  
 KEYWORDS  
 SOURCE Streptomyces nanchangensis  
 ORGANISM Streptomyces nanchangensis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 132544)  
 AUTHORS Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
 TITLE A Complete Gene Cluster from Streptomyces nanchangensis NS3226  
 Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin  
 Chem. Biol. 10, 431-441 (2003)  
 JOURNAL 2 (bases 1 to 132544)  
 REFERENCE Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
 AUTHORS Direct Submission  
 TITLE Submitted (13-JUN-2002) Bio-X Life Science Research Center,  
 JOURNAL Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
 China  
 FEATURES Location/Qualifiers  
 1..132544

```
/organism="Streptomyces nanchangensis"
/mol_type="genomic DNA"
/strain="NS3226"
/db_xref="taxon:204925"
complement(682..2052)
/gene="nanT1"
/gene="nanT1"
/feature="putative integral membrane transport protein"
/codon_start=1
/transl_table=11
/product="NanT1"
/protein_id="AAP42848.1"
/db_xref="GI:31044136"
/translation="MHOTTDDPRASRRSGVRRFAVSLTGTALEWYDFAVYAAAA
LVFGDLFPSPEDPLTGLLAPFVAGVVSRLPGIIVFGRIGDVIGRKKVLIATLVI
GAATFLGLVPTSTIGVAPIALVIRFAGGVGVGEGWGAVALLSSEFGDSRRRGFY
ASAAQVGGPNGLANGVLAAGALLTEAQFEAWGRVAFLLSGALVAFGLIRAKLE
ETPVKAEAPQSRPEAPIREVFTTQPRALLAAILCRVPDVLVAFVFTVLTATGE
LMSRGALAAVLIGSSIQVFLMPLAGAVSDIRNRVLYGCAVAAVGMPLFPFMIG
GSMIPLAGVVGVLVTHSFYGPAAFIABQFSPRLRYTGSSLAYTLAGIIGAIAP
LFTLLSAGSWLPLALYIAVAASVLAGVLLGRDPEAAVDEDAQLATPAQAQATA
ARPTAV"
2064..2879
/gene="nanT2"
2064..2879
/gene="nanT2"
/feature="putative ABC transporter"
/codon_start=1
/transl_table=11
/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translation="WLGHRPSGILLVYRSYAQALRKLTAEKSTTADRHNSGAPM
ARLTFELPDGSTVDIVQVNLNAGYAGSQQDDVAHIAELAEGLVPTSPYALYPA
PYLAQIDRVAGVQRRSTGGEAEWALVAGDGELLTACDHTDRDLEVHGVAWKNAG
PDVLARWLADVEPRDLDTLRWVTRDGTETEIQHTLAELLTPAYWVVLRSRD
ALTPTGLISGTIPMTPGVDQFADTWVRVELGDPATGDTIRLAYDVHPMPEPIG"
3117..3818
/gene="nanT3"
3117..3818
/gene="nanT3"
/feature="putative two-component response regulator"
/codon_start=1
/transl_table=11
/product="NanT3"
/protein_id="AAP42850.1"
/db_xref="GI:31044138"
/translation="MSTRCTGSLTNGWVSVELGIPLSVVLVDDHPVVRAGISAWCAA
DAPISVVAEGANVSVALHGPGRGADVVVMDILLQNGRPAYDELQELVAQERKVVYTM
RDSQDAALTCMDLGSATYITKAGQRLVKAIRAAEDIPYTPSLAGAFSGDTRQSR
PVLVSREVELVWFQSESKAVVAQSLGISERTVNTYLDVRVRIKYANAGRPAATTKAKL
VARAVQGLIALDEL"
complement(3724..4923)
/gene="nanT4"
complement(3724..4923)
/gene="nanT4"
/feature="hypothetical chemoreceptor"
/codon_start=1
/transl_table=11
/product="NanT4"
/protein_id="AAP42851.1"
/db_xref="GI:31044139"
/translation="WGEDFROLITPHQLRIGRAGGLVLTVLTHIGSALPGLIGLGEY
RHPPIPLAAYGLLTLILGSSVVLGDRPWRGWPVPCALAGTFVASVATTQQVPADHY
FVSSLHWSYSLAGNVTLLGSRPLVSGAFGLAHVLAATLTLVDPVPSISGASMA
LSASAGCFQMTIRAGAKLLDSSAICQALRAERVTRIAVAFIQDQRRYAEI
NATNLTLLGLASLDDEDDHACALEARLRFAESDCTFDPLVHEMRACIDV
ABRNQTVQLAVRGDPLEPLPLRLRLDLPVLAALAAAGQARTVTVRGDQVRGVV
VDALDRDPEPKANGIRVTRVHAEDRLVVEACRITARPESLDRPEPPWPAPWSWP
AGRSRT"
complement(4916..6106)
/gene="nanT5"

/organism="Streptomyces nanchangensis"
/mol_type="genomic DNA"
/strain="NS3226"
/db_xref="taxon:204925"
complement(682..2052)
/gene="nanT1"
/gene="nanT1"
/feature="putative integral membrane transport protein"
/codon_start=1
/transl_table=11
/product="NanT1"
/protein_id="AAP42848.1"
/db_xref="GI:31044136"
/translation="MHOTTDDPRASRRSGVRRFAVSLTGTALEWYDFAVYAAAA
LVFGDLFPSPEDPLTGLLAPFVAGVVSRLPGIIVFGRIGDVIGRKKVLIATLVI
GAATFLGLVPTSTIGVAPIALVIRFAGGVGVGEGWGAVALLSSEFGDSRRRGFY
ASAAQVGGPNGLANGVLAAGALLTEAQFEAWGRVAFLLSGALVAFGLIRAKLE
ETPVKAEAPQSRPEAPIREVFTTQPRALLAAILCRVPDVLVAFVFTVLTATGE
LMSRGALAAVLIGSSIQVFLMPLAGAVSDIRNRVLYGCAVAAVGMPLFPFMIG
GSMIPLAGVVGVLVTHSFYGPAAFIABQFSPRLRYTGSSLAYTLAGIIGAIAP
LFTLLSAGSWLPLALYIAVAASVLAGVLLGRDPEAAVDEDAQLATPAQAQATA
ARPTAV"
2064..2879
/gene="nanT2"
2064..2879
/gene="nanT2"
/feature="putative ABC transporter"
/codon_start=1
/transl_table=11
/product="NanT2"
/protein_id="AAP42849.1"
/db_xref="GI:31044137"
/translation="WLGHRPSGILLVYRSYAQALRKLTAEKSTTADRHNSGAPM
ARLTFELPDGSTVDIVQVNLNAGYAGSQQDDVAHIAELAEGLVPTSPYALYPA
PYLAQIDRVAGVQRRSTGGEAEWALVAGDGELLTACDHTDRDLEVHGVAWKNAG
PDVLARWLADVEPRDLDTLRWVTRDGTETEIQHTLAELLTPAYWVVLRSRD
ALTPTGLISGTIPMTPGVDQFADTWVRVELGDPATGDTIRLAYDVHPMPEPIG"
3117..3818
/gene="nanT3"
3117..3818
/gene="nanT3"
/feature="putative two-component response regulator"
/codon_start=1
/transl_table=11
/product="NanT3"
/protein_id="AAP42850.1"
/db_xref="GI:31044138"
/translation="MSTRCTGSLTNGWVSVELGIPLSVVLVDDHPVVRAGISAWCAA
DAPISVVAEGANVSVALHGPGRGADVVVMDILLQNGRPAYDELQELVAQERKVVYTM
RDSQDAALTCMDLGSATYITKAGQRLVKAIRAAEDIPYTPSLAGAFSGDTRQSR
PVLVSREVELVWFQSESKAVVAQSLGISERTVNTYLDVRVRIKYANAGRPAATTKAKL
VARAVQGLIALDEL"
complement(3724..4923)
/gene="nanT4"
complement(3724..4923)
/gene="nanT4"
/feature="hypothetical chemoreceptor"
/codon_start=1
/transl_table=11
/product="NanT4"
/protein_id="AAP42851.1"
/db_xref="GI:31044139"
/translation="WGEDFROLITPHQLRIGRAGGLVLTVLTHIGSALPGLIGLGEY
RHPPIPLAAYGLLTLILGSSVVLGDRPWRGWPVPCALAGTFVASVATTQQVPADHY
FVSSLHWSYSLAGNVTLLGSRPLVSGAFGLAHVLAATLTLVDPVPSISGASMA
LSASAGCFQMTIRAGAKLLDSSAICQALRAERVTRIAVAFIQDQRRYAEI
NATNLTLLGLASLDDEDDHACALEARLRFAESDCTFDPLVHEMRACIDV
ABRNQTVQLAVRGDPLEPLPLRLRLDLPVLAALAAAGQARTVTVRGDQVRGVV
VDALDRDPEPKANGIRVTRVHAEDRLVVEACRITARPESLDRPEPPWPAPWSWP
AGRSRT"
complement(4916..6106)
/gene="nanT5"

complement(4916..6106)
/gene="nanT5"
/feature="putative two-component sensor histidine kinase"
/codon_start=1
/transl_table=11
/product="NanT5"
/protein_id="AAP42852.1"
/db_xref="GI:31044140"
/translation="MSISSLSPGIDPPASARLPGSAGPWAREYVAFNRRLIYGRAGV
MSCAVLGVAMPWEMGAPAAVAAVLAWSWHLRLATTSAMQRLTLLALDLTMTA
LCLSQRTIPEAQTHAGTILYVSFTAVAYQLMQPLTGILLATLCCADVVGTA
HPNLESGVLRSVIWWNTALARAVALVVSFESRAADRAARAARQOGEAEARAC
AAERHIAALHDTACATLLIASAPWRAETLRAQAARDLLRURASQPVAGEVLLAA
ELDDEIAHLPLRINRFDELVTTRPIAALRGGLGEALRNVARVAGVADVARTABR
VDHVIVTTIADSGVGFDEPHIPGHRIGKRSIHGRMNVVGVRATVESRPGHGTTRLE
WPRG"
complement(6768..7493)
/gene="nanR1"
complement(6768..7493)
/gene="nanR1"
/feature="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR1"
/protein_id="AAP42853.1"
/db_xref="GI:31044141"
/translation="WGPLRVVDERKSGSPRAKLAALVAVLLIRSDOVVMAEQLTREI
WGPLPRRDYASIVVCI SRLKFLSRPQQNPITRPGYMLRRASCECDPFDVQQLV
ADGKCYVGEPAQAALCFEQALAEWGPAGNVQTCPIVAGFLKRLDEMRECAEMS
AAHLELGRHRELVRSLRSLVAEPFLRETTFYRQLMALYRSDQSEALGIYESARRVL
DRELGEPGRKRLRELHRTVLAST"
complement(7573..8334)
/gene="nanR2"
complement(7573..8334)
/gene="nanR2"
/feature="putative regulatory protein"
/codon_start=1
/transl_table=11
/product="NanR2"
/protein_id="AAP42854.1"
/db_xref="GI:31044142"
/translation="MRYEILGPLVDRSSVTSISARKIEVLVLAALLVRADVVSSD
LVKEIMTRAPQRAALHYVYQRLKFLRPNETSPITVARSYVLRIGDDELDFH
VQGLVRGAAQAGRAQTHAEACEAYEEGFWEFGFVLEDLFDGAIVNGFVCAEQARL
ECVFSFVAGLELGRHEFVSYLYARIEELPLNEAFYRQLMALYRSCRADALHYQ
SARAVLSEELGLEPSAALKRLHHAVALVDAPELEAA"
8919..17627
/gene="nanA1"
8919..17627
/gene="nanA1"
/feature="PKS"
/codon_start=1
/transl_table=11
/product="NanA1"
/protein_id="AAP42855.1"
/db_xref="GI:31044143"
/translation="MRFYGVWAGSSPSHAQATSPVAVLGLACRLFGAPDPEAFWR
LLRAGENAVVPDSRLPTEGPPYFAGLLEHVDITDGFPGFISPREATMDPQORL
MLELWAMEDAGLGPKNLAERTAVFTGATWDDYATLLHRRRNDIALTRHTWAGLR
RGLIANRVSHLLGRGSLTVDAAQSSGLVAVHACESLREGEADLALAGGVNLLAE
ESMMAEQFGLSPDGKCYTFDARANGFVGGGWWLLPLAAAVADGDPVVCVIE
GSVNDGATDGLTRPSADQTDVVRQAWRAVSPAELOVLEHGTGTVPVGDPIEAA
ALGDALGDGCHRDRAVDAPLRVGVKTVNGHLEAAAGIAGLTKLSTIHRRLPP
SLNFTAPGIPLAELGLRVCTAFGPPDERRLTAGVSPFGMGTNCHVVLAPPPAP
AVLPDRTPAGDVSARTAPPVMPVVSASPALKTAQAAALYELHRAHPGLHPVDIGH
ALATRTAFRAVAVVLDREDELVSRDLALASETQTSVINGRAGGRVAFILFSGQS
QRPMGRLAAAPVVFADALREKCHLDMLDTPDPLDLMFAAPPDGPPLNKIAY
TPQALFAIYVALYLVTSWGTDPHLMHSGVGEITAAHVAAGVLSLPACTIVARGRL
MOSITAPCAMAAQWATAEAGQALAEYGVGLAAVNPASVITSGOREVAEATAAW
PARGKATVLRKSHAFHSLDGLTDLRTVAAGLTAAPAI PVVSNLTGGAATEAOL
RSPDIYADHARQAVRFADVAGVCDAGVDTLELGPDASLTGASLRASNAAGAPRP
VAVAVQRGRPEAQSFVSMAHQHVGVDWAAAFAGHETRRAPLPTIYAFQRDRHWP
DGLDERGARREPTSPVVFESTDREPVVVDASPADRAASPGELLALVTRTHAALVGHNSP
```

DGIDPALTFFKQGFDSLAATLSERLSAATDTTELPATLTFDHPPTNVAWAALRAAHEG  
QPTAATAAGPESMAEDPVAVAVSCRYPGVGESEALWRLVDEGVDAVGFPPDGRW  
DLAELFGAPDGSATGCGFLYAGAGDFDAEPFGLSPREALAMDPOORILLLELSWE  
LLERAGIIPASLASATGVYVCAATDYGPRLHEATAELODGHLLTGSPSVASGRVAY  
ALGUEGPAITVDTTSCSSLVAMHIAAQLRQCECDLALAGVQVMTATGPMFTSFRQR  
GLAPDGPCKPAAADGTGMSSEGAGLVLLERLSDARRNGHQVLAVIRGSAVNQDGSN  
GLSPNGPFSQORVITQALANARLPADVDVAEHAHTGTGTLGDPTEAQAALLATYGCORT  
DDRLPLWLSIKNSITGHTQAAAGVAVIKMVALRGLRPLASHIDAPSHIDWSDGTV  
RLLESPDMPCTDMPGSDRPRAAVSFGISGTNAHLILEOAPDHPPEPTTSGGVTV  
WYLSARTADALRAQAAGLAEWWTAGAPRSPASPTSPADYGVGSLATRSADRHRV  
VSGTRDELLEGLRAVADGLAPAVSAGAAFGPVVFPQGSQWRGMSVELLSSPVF  
AARMAACEALGSLRATLVRGAPGAPESRVDVLOPCLWAVMVSLLAVWESYGV  
TFTVADVSGOGEI1AAACVAGSLRDCGAVVVALRQAALRAGHGTWASLALSGAEAE  
RFLADLGAARVAVVAVPGYSTVVSQPTDQVAAVAAACEAGHRAARTIDVYASHG  
POVRLADTITDLDLSPGASDAVYSAVTEGARQTEELDADYFNTLROPVFASA  
IDALLAGYRVFIEVSHPPVLLPALRECFEEAEVAATVPTLRDQGGDPQVAAALGD  
GVAGLAWDRGVRFGDREAGDGHRTVELPTYPQRRRYWLPADPHGRREGRTAG  
VGTTPAGHALSSAVELADGGLVLSGRLPFGDAAGVAGVQVLPVGAVIDWALLA  
ADEAGASLELLRLRAPELLELSPGLSGLSGLAVQAAGVAGPDSGRRELRSRPA  
AGAGEWTCVAVGSLAPGPPAPADTGTATVPWPPAGAEADPAGLYERARRGVYGV  
PALRGVVALWDGADLVADVALPREAGGGGEGGADGCTAGFGHLPVLLDALOPALI  
AEPDGTGEGGAPARLWLPFWAGSVRLMATGARARVRLSPDGGGGDVADERELRI  
EVSQDTPVLUSVSVLRPTVRQVREASGAAGGLFALDWTVPVAPPSGAEDDAG  
CVVLGEAPTEFGVDCSDTITDLPALIALDAGAPLPSVVMVPPAADGCAEDDAA  
LSAVRGVAAALRAVWAEPLTVSLVAVTRGAAAGGAEGFPVLDLAAAGWCAARGVQ  
AEHPDRIVLVDVDDVDMDADTDDIGAAAGLAAAGLPQVALRGDTLLAPLARSAA  
TPGGVAFDPTGTVLVDSDGSLAGSVAEHLVRAEGVRLHLLVPRREGAGVAYTDVROD  
AQVHWTVDPDRTAALERVAQVDPAPHLITGVVHVAGLSADIESGARGHVAAGVVA  
RALHQATLASFVRLTUSDRATWDGPAAPERAAGAFCAATVDVRRRAGLHGLDVA  
FGPWAADDDGAGSGGRTVGVIGADRGALLRACRADRPRLVAAGDITRITALTAHPA  
HELPAALRTGASASASAGRAPVRVRAAAAPGRTTWDWASRLVGLGPAERBRVLELV

Query Match 5.2%; Score 26; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AGGCGCTGCTGCCACCTACGCCAC 468  
|||||  
Db 75927 AGGCGCTGCTGCCACCTACGCCAC 75902  
|||||

RESULT 40  
SBEERYA  
LOCUS SBEERYA 9769 bp DNA linear BCT 14-OCT-1991  
DEFINITION S. erythraea eryA gene for 6-deoxyerythronolide B synthase.  
ACCESSION X56107  
VERSION X56107.1 GI:46975  
KEYWORDS eryA gene; erythromycin.  
SOURCE Saccharopolyspora erythraea  
ORGANISM Bacteria; Actinobacteria; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.

REFERENCE  
1 Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F.  
AUTHORS An unusually large multifunctional polypeptide in the  
TITLE erythromycin-producing polyketide synthase of Saccharopolyspora  
erythraea  
JOURNAL Nature 348 (6297), 176-178 (1990)  
MEDLINE 91043075  
PUBMED 2234082  
REFERENCE 2 (bases 1 to 9769)  
AUTHORS Leadlay, P.F.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1990) Leadlay P.F., Dept. of Biochemistry,  
UNIVERSITY of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK  
LOCATION/Qualifiers  
FEATURES  
1. .9769  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL 2338 (sub-species)."   
/db\_xref="taxon:1836"  
57. .9593  
/gene="eryA"

mat\_peptide 57. .9590  
/gene="eryA"  
/product="6-deoxyerythronolide B synthase"

ORIGIN  
Query Match 5.0%; Score 25; DB 1; Length 9769;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CDS

57. .9593  
/gene="eryA"  
/codon\_start=1  
/transl\_table=11  
/product="6-deoxyerythronolide B synthase"  
/protein\_id="CAAJ9583.1"  
/db\_xref="GI:46976"  
/db\_xref="GOA:Q03133"  
/translation="MSGDNGTEELRLRYLKYRTVELDSTVARTLEVEVHAGEPTAIV  
GMACTPVDVDPSPSFMFVSGGDAIAEAPADRWEDDPDARLGMLAAGDDPAGF  
FGISPRALAMDPOORIMLEISWEALRAGHDVPSLRGATSGVFTGVTVDGPPDE  
APDLVGTGTGTSVASGRVAYCLGLEGPAMTVDATSCSGLTALHILAMESLRDEC  
GLALAGVTVMSGTAFTFVSQGLAAGDCKPFSKAADQFGLAEGAGVLVQLRSLA  
ARRGRPVLAVLGASVAVNDCASNGLTAPSPAQORVIRRALENAGVAGVDDVYEAH  
GTGTLEDPDIEVHALSTYGABERDPPDLWTGYSVKSNIHTQAAAGVAGVAKVIALR  
HGMETLHFDPSPQIEMDLCASVSVQASRWPAGERPRRAGSVSGISGTHAVIV  
EEAPEWDEPPADSPGVPVLVLSGRDEQMAQAGRLADHLAASGTRCATSRWPP  
AAAEQVQVGRDEALAGLAVADRITADRTATGQGNPSRRVAMVFPQGAQWQM  
ARDLRSQVPAFDSIRDCERALAPHDMSLTDLISGARPLDRVDVQVPLFVWVSLA  
ALWESHGVEPAAVVHGSQETAAARVAGALTLEDAAKLVAVRSRVLRLGGQGMASF  
CLGTEQAARIGRFAGALSIASVNGPSSVVVAGSGPLDELALACEAEHAKRIPV  
DYASHPSRELELLTELAGISPVADVALITTTGQPIDTADMTATVYANLREQ  
VRFQDQTLAERAGDAFVPSHPVLTVGLIATLDSALPADACVATYTRDRGGL  
ADFTALGAAYQGVVDWSPAFADARVELVFPFQRELVPYFQRYQWLPITGG  
RARDDEDMRYQVWREAEWESLAGRLAVLTVGPVPSLSDALRLGSEQSGATVLT  
CDVESRSTIGTALRAADTALSTGVAAVPHGEAVDPSLDLALVQALGAAGVAPLW  
VLTRNSVQAGELVDPAQMVGGLGVVGLQEPGRWGSLVDLVDADAASIRSLAIVL  
ADPRGEBQVIRADGIVKVARLVAPARATHEPLEPLAGTVLTGSGIGHALARSQ  
AEHLVILGRGADAPGASSELREELTAGTGVTTAAACDVADRARLEAVLAEEAABGR  
VSAMVHAAVSTSTPLDILTAEETIADVVRGTVDNDELCPDLDAFLVSSNAGW  
GSPGLASVAAANFLDGFARARSEGAPVTIANGLWAGMAGDEGEYLRARQSE  
MDPRVABELHITLDGQTSVVDMDRRRVELFTAAHRPLFDEIAGARAEARQSE  
EGPALAQRLLALCDGRERHHLIRAEVAALVGHGDDAALDRDRARLDFGDSMTA  
VDLNRLLAATVGRAAATVFDHPITRLADHYLEELVGAEEAEPALRVREVPKDA  
DDPTAIVGMACFPFGVHNPGLMEFTVGGGDAVTEMPTDRGWDLDLALDPDPRHGT  
DYSRHFAGLDAAGDAFAAFGSGPREALAMDPPQVLETTWELFENAGLDPHSVRGS  
DTYRGLGAAYQGVQDAPVEDSEGLITGNSAVSVSVAYVLELEGPAVTVTACS  
SSILVALHSAACGLSDGCLGAVAGGVSMAGVEVFTFSROGLGADVCRCKAFSAED  
FGFLPEGVAVVQLQRLSDGPAEGGQVGVVAGSAINQDGTNGLAAPSGVAQQRVIR  
KAWAPAGITGADVAVEAHGTGRLGDVPEASALLATYKSKSGSGPVLGSSVKNIG  
HAQAAAGVAGVVKVILNRLGLVPPMLCGRSPSLIWSGGVSELAEEAVPMPAADG  
VRAGVAGSVGSGTNAHVIIIAEPPEPELPPEFPGVGLAANSVQLLSARTETLAA  
QARLLESVADDSVPLTALASALATGRALPRRAALLAGDHEOLQVLSVAGFSAEVL  
ATTGTASAGGVVFPVPGQAGWEGMARGLLSVVFAESIAECDVLSVAGFSAEVL  
EQRPDAPSLERVDVQVPLFVWVSLARLWCAAGVSPSAVIGHSGIETAAAVVAGVLS  
LEDGVRVVALRAKARLALAGKGMVSLAAPGERARALITAPWEDRISVAANSPPSVV  
SGDPEALAEVARCDEGVRKTLVDYASHSRHVEEIRETILADLDGILARRAIIPL  
ADAVAIGSLHRDTABEHLIAELARAHVGHVAVDMENVPAPPPVAPLPYFPFQRYW  
APEVSDOLADSRVYDWRPLATTVDLGGFLVHGSAPESLTSVAKGAGVVPVASAD  
REALAALREVGEVAGVLSVHTGAANALHQSIGEAGVRLAPLSRAVALGESE  
PVDPEQAMVGLGRVNGLETPERWGLVDLPAPGAPGGEAPVACLGDGHEDQVDIR  
DHARYRLRVLAPLRTRESSWEPAGTALVTGTGALGHVHARHLCAGVEDVILVSR  
GVDPAPAAAELEIIVAGLFTTITACDVADREQLSKLEELRQGRPVTVVHTVGP  
ESRPLHTEIGELSVCAKVTAKVLDLCPDAETFLFSSGAGVWGSANGLASANA  
YLDALHRAVRKAVERTSVAMWAGMAGMTGLEGLTFRGLRPMAPORATRALHQAL  
DNGDCTVSIADVDMEEFAVGTAPRPLDDELVTTPAVGAPVAVCAAPAREMTSEL  
EFTVSHVAALIGHSDGPDVAGQDPTTELGFSDLSLTAVALRNQLQATGLALPATLVEH  
PTVRRILADHIGQSDSGTPAREASALDRQVAGSVGRSVRSYLDLGLSDFRHHF  
GSDGFLDLVDMADGEGVTTCAGATAAISGHEFTRLAGLRGIAFPVAVYPPQGYE  
EGEPLSPSMAAFAVQADAVTTCQDKPFPVAGHAGALMAYALATELIDRHPRPGY  
VLIDVYPPQHODAMNMLELTATLFDRETVRMDTDLTALGAYDRLTGQWRPRTGL  
PTLLVSAGEPMPGPPDDSNKKTWPEHDTVAVPGDHFVQVEHADAIARHIDAWLGGG  
NS"

QY 329 CCGACGCGCGTGCATCCGACGGC 353  
 |||||  
 Db 5389 CCGACGCGCGTGCATCCGACGGC 5413  
 |||||

RESULT 41  
 AF220951  
 DEFINITION Streptomyces antibioticus 8,8a-deoxyoleandolid synthase 1 (oleA1)  
 and 8,8a-deoxyoleandolid synthase 2 (oleAII) genes, complete cds.  
 ACCESSION AF220951  
 VERSION AF220951.1 GI:9049534  
 KEYWORDS  
 SOURCE Streptomyces antibioticus  
 ORGANISM Streptomyces antibioticus  
 Bacteria; Actinobacteria; Actinomycetales; Streptomyces.  
 1 (bases 1 to 24568)  
 Shah, S., Xue, Q., Tang, L., Carney, J.R., Betlach, M. and McDaniel, R.  
 Cloning, characterization and heterologous expression of a  
 polyketide synthase and P-450 oxidase involved in the biosynthesis  
 of the antibiotic oleandomycin  
 J. Antibiot. 53 (5), 502-508 (2000)  
 20363406  
 10908114

REFERENCE 1 (bases 1 to 24568)  
 Shah, S., Betlach, M. and McDaniel, R.  
 Direct Submission  
 Submitted (04-JAN-2000) Korean Biosciences, Inc., 3832 Bay Center  
 Place, Hayward, CA 94545, USA  
 JOURNAL Location/Qualifiers  
 MEDLINE 1..24568  
 PUBMED /organism="Streptomyces antibioticus"  
 AUTHORS /mol\_type="genomic DNA"  
 TITLES /db\_xref="taxon:1890"  
 JOURNAL 623..13075  
 623..13075  
 /gene="oleA1"  
 /note="polyketide synthase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="8,8a-deoxyoleandolid synthase 1"  
 /protein\_id="AAF82408.1"  
 /db\_xref="GI:9049535"  
 /translation="MHVPGRENGHSIAIVGIACRLPGSATPQEFWRLLADSADALDEP  
 PAGRPFGSLSSPAPRGGLDSIDTFADFNISPREAGVIDPQORLALBELWEALE  
 DAGIVPRHLGRTRTSVFMGAWDDYAHLAHARGEALTRHSLTGTGRGMIANRISYAL  
 GLQPSLFTVDGTGSSSLAAVHMACESSLARGESDLALGVGNLVDPPAGTTGVERFGL  
 SPDGRCTYDFSRANGYARGEGVVVVLKPTHRALADGTVYCEILGSALNMGATGEGL  
 TVPSARQADVLROAWERARVAPTDVQVELHGTGTAGDPVEAEGLTALGTARPAE  
 APLLVGSVKTNIHLEGAAGTAGLTKVLSIKRHLPAISNTPSPNPRIDLDALRLV  
 HTAGPSPSPRPLVAGVSFGMGTCNVHLSRLNAGGCGAGKPYGTGTELRIGAT  
 EAKRPPDPAIGDPPADQTHRYPLIILSARSDAALRAQAEIRHHLHSEHSPGLRDT  
 AYSLATRQVFERHVVTHGRDLENGRLDLENGLPAPQVLLGRTPTTPEPGGLAFLF  
 SQGQSQPMGRKHLQVPPFRDALDEVCSELTHGLRGLPEAGPPLRDVWFAERGT  
 AHSALLSTHYTOALFALETALFRLVWGLKPDHLAGSVGHTAAAHAGILDLSD  
 AAEIVATRGALMBSLPGGVMSLQVQAPESVAPILLGREAVGLAANGPDAVVSGE  
 RGHVAALEQLDRGRKRYLRVSHAFSPMLPEVLEFAEAVAGLTFRATPLVSN  
 LTGAPVDRTWATPAVVRHREAVRFGDGRALGKLTGSFLEVPDGVGLTNMARAC  
 VTAAPGHRGQGAADATALLPALRRGRDEARSLTEAVARLHLHGVPMDTNSVL  
 GGDVSRVPLPTAYFQRESHWLPSPGAEPRPADTESGTGTAEPSPHVDVLRVSH  
 AAVALGHRAERI DPDRALDGLFDSLTALELRDLRDLTALGLRPLSSVLPDHPSPGAL  
 ARELQGDTRPEPKTKNGTRATGPDDEPTAIVGMACRFPFGVTSPPDLWRLLA  
 AGEADVSPFPDGRNVYDSINRGGFLYDAGEFDFAFGISPREALVMDPQKLLLE  
 TSWALERAGVSPGSLRSDDTAIVGATQDYGRLHESDDSGYVLTGNTASVAG  
 RIAYSLGLEGPVTVDTACSSSLVALHLAVALQARRGECSLAGAGATWMPSPGAVFV  
 SROGLSDEGDKAFATADGTGWAEGVLLVERLSDRILGHRHLAVVRGSAVNQD  
 GASNLGTAPNGPSQORVIRALADAGLVPADVVEAHGTCGTGLDPTLEAALLATYG  
 QGRAGRPVTLGVSVKSNIGHTQAAGVAGVMKVLALGRGVVPTLHVDERSAHVDWS  
 AGEVELAVEAVPWRGGRVRAAGVSFGISGNTAHVVEAPAPSEPEPEPGRGVSVG  
 VVPVWSGRDAGALREQAARLAHVGVSVDVGVMSLVATRSVFHRVAVMVMSGELDAM

FEATURES  
 source  
 misc\_feature  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS-Q"  
 2309..3415  
 /gene="oleA1"  
 /note="Region: acyltransferase domain AT-L"  
 3485..3724  
 /gene="oleA1"  
 /note="Region: acyl carrier protein domain ACP-L"  
 3806..5056  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS1"  
 5363..6400  
 /gene="oleA1"  
 /note="Region: acyltransferase domain ATI"  
 7109..7669  
 /gene="oleA1"  
 /note="Region: ketoacyl reductase domain KRI"  
 7943..8200  
 /gene="oleA1"  
 /note="Region: acyl carrier protein domain ACP1"  
 8258..9541  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS2"  
 9848..11082  
 /gene="oleA1"  
 /note="Region: acyltransferase domain AT2"  
 11723..12274

AESLAGFAAGGVVPGVWPAEGRRVVFVPGOGSQWGMAGLGLDACPVEAFAVA  
 ECAVLDPGLTGNLSLEVLGGGEAVLGRVDVVQPALWAMVSLATWYGYVEBRAVAV  
 GHSQGEIAAACVAGGLSADGARGVLLRSRALRATAGGGGMSVSLPGRVMTLLEEF  
 DGRVSAAVNGPSSVSGDVOALDELLAGCEREGVRRVVDVYFSAQMDQLRDD  
 LLEALATVPTSSANVFFSTVTDALDAGYFNTLRETFRFQBAVEGLVQAGM  
 GAFVCSPPHVLPTITETLDFDADAVALSRLRDEGLDFLTSLAEAFVQGVVD  
 WSRAPEGASPTVLDPTVFFORWLLDKAAERERLEREDWRVHVRPVTTPRSAR  
 LSGWVAIPARLARDSLLVGAIDALEGGARAVPVVDERDHQALVEALRNLGLD  
 DDLGVALSLILDENPHGDHDPVPGMAASLALVQAMADAAEVEVFWATGAAALP  
 GESPERQALLNGLRVVALLEQPTWGLVLDLPOHLEDAGRLVVDVGGADBDQL  
 GESSPVLARLVRTPGHRMSQAGREWSPTVLTGTGTCALGAVHARMLAGAE  
 HLVTISRGADAGAAALSLDITDMGVRTLAACDADHALETLLDLSLTDPAQTLA  
 VTHAAGLDGKMTVLTPPEOMNALRAKVTATVNLHETRLDILDSAFVLSISATLG  
 IPGOANYAPGNSFLDPAEARRAAGLAVTSIANGPWSGGTGMAGHSVGERLQRHVL  
 AMEPAAIALDHTLASDETAVAVADIDWSRFFLAYTALRAPLIGTPEAKRMLES  
 SGPDLEPRAPLAVLAGLTAVEQLRLVQLVREQAAVVLGHSBAEAVAPDAFK  
 DLGDSLTSVELNRNLATGLRPLVTFVAFYARPAALAGHLSLIDDDGHALPG  
 VERHAIDEPFIAIVGMACFPFGIASEPDLMDVLTAGEDVWSGLPNCRWDLRLYD  
 PDRAGTSYMRGAFHEAGEFDAFFGI SPREALAMDPOORLLETSEWALERAGITP  
 SKLAGSPGVFFGMSNODYAAQAGDVPSLSEGLYLLTGSISVASGRVATYTGLEGPAV  
 TVDTACSSSLVALHLAVOGLRGCSSLALGVGVTVMSRFTLITPSRQGLSDGRCK  
 APASADGFGAAGGVVLLVERLSDRLEHRLVAVRGSVAVNOGASNGLAAPGNS  
 QORVIRALADAGLAPADVVEAHGTCTRLGDTLEAQLLATYCGRTSGRPVWLG  
 VKSNIGHTQAAAGVAGVMKVLHGVVPTKTHVDEPSHPVDSAGVELAVEAPVW  
 SRGRRRAGVSFGISGNTAHVVEAPAPSEVEEGSPGVVGVVWVSGRDAGALR  
 AQAAARLAHVSSTGAGVVDGMSLVATRSVFHRVAVMGTDLDSWAGSLAGAGV  
 PGVSGVAPASGRVVPFGQSGQWGMAGLGLDACPVEAFAVACCAALDLRTGWS  
 LYEVLRGEAVLGRVDVVQPALWAMVSLATWYGYVEPAVVGSGEIAAACVAG  
 GLSADGARGVLLRSRALRATAGGGGMSVSLSAERTMLDITGGRVSVAALRQSS  
 TVSGDQALDELLAGCEREGVRRVVDVYFSAQMDQLDELLALADVTPODSS  
 VPFSTVTDALDAGYFNTLRETFRFQBAVEGLVQAGMAGFVCEPSPHVLV  
 GITLDTDFDADAVALSRLRDEGLDFLTSLAEAFVQGVVDWTHAFEGRPEFVD  
 LPTAFORORYLHEEPLOEPVDENADAEFWSVVERGATATVSLDSTDAALHTVLP  
 ALSWRRRVRRHLQWRVYEWKPPPALELDVILGGLVFPFRLGDDHGVVARKVA  
 AVTARGEVSVVELDPTGPRRAVAAVAGVSGVSVFSLSDRRRSEHSEVFPAGLA  
 ASLVLAQALVDLGRGEPRLVTRGAVAGSPDAGVVIDPQAVQVWGFGRVGLGLEH  
 PELMGVLVDLPVGDVEEVRFFVGVASAGPEDOVAVRGSVWVRRLVRAVVDGGGG  
 WRPRGTVLVTGGLGGLGHTARVLVGGGADHVLVSRGGSAPGAGDLVRELGGA  
 RVSVRACDADRVARLALLSDLGEPVTAFAAGVPOSTPLAEISVQBAADVMAKVA  
 GAVNLGELVDPGLEAFVLFSSNAGVWGSQGAAYAAAFDALAVRVRGVGUPATS  
 VAWMGAGENASVGGAAARLSRRGRVAMDPRAVAVMADVAGEAFVAVADVWER  
 FVTGASARPREPLISDLPEVAVVEGVQGRGQGLGVGEESKRLKRLSGLSRVQ  
 EELVELVRAQAAVVLGHSADVPABAFKELGFDLSLTAVELRGLAAATGILRPT  
 MAFDPTATATARFLQSELVSGDPLTMRSAIDQLETGLLLESDSEARSEITKELN  
 ILLPRFGSGSSRRGAGQDAGEQDVEDATIDELFEVLDNELGNS"  
 650..1906  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS-Q"  
 2309..3415  
 /gene="oleA1"  
 /note="Region: acyltransferase domain AT-L"  
 3485..3724  
 /gene="oleA1"  
 /note="Region: acyl carrier protein domain ACP-L"  
 3806..5056  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS1"  
 5363..6400  
 /gene="oleA1"  
 /note="Region: acyltransferase domain ATI"  
 7109..7669  
 /gene="oleA1"  
 /note="Region: ketoacyl reductase domain KRI"  
 7943..8200  
 /gene="oleA1"  
 /note="Region: acyl carrier protein domain ACP1"  
 8258..9541  
 /gene="oleA1"  
 /note="Region: ketoacyl synthase domain KS2"  
 9848..11082  
 /gene="oleA1"  
 /note="Region: acyltransferase domain AT2"  
 11723..12274











GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgtctacacgtcgggcac.....cccggtcgagggctcgtcg 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hcg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_un.\*

28: em\_vl.\*

29: em\_vi.\*

30: em\_hcg\_hum.\*

31: em\_hcg\_inv.\*

32: em\_hcg\_mus.\*

33: em\_hcg\_oth.\*

34: em\_hcg\_pln.\*

35: em\_hcg\_rod.\*

36: em\_hcg\_nam.\*

37: em\_hcg\_vrt.\*

38: em\_sy.\*

39: em\_higo\_hum.\*

40: em\_higo\_mus.\*

41: em\_higo\_oth.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
C 4	302.6	60.4	82746	1	AF453501	AF453501 Actinosyn
5	285	56.9	69644	1	AY179507	AY179507 Streptomy
C 6	275.4	55.0	32870	1	AF007101	AF007101 Streptomy
C 7	243.8	48.7	2614	1	ATTECR	X56928 A. teichomyc
C 8	243.8	48.7	2614	6	A21096	A21096 Synthetic D
C 9	192.2	38.4	107379	1	SHGCP1R	X86780 S. hygroscop
C 10	184.2	36.8	28732	1	AF082100	AF082100 Streptomy
C 11	184.2	36.8	77534	1	AF235504	AF235504 Streptomy
C 12	184.2	36.8	77536	6	BD235937	BD235937 Polyketid
C 13	184.2	36.8	77536	6	AR271638	AR271638 Sequence
C 14	163.8	32.7	77457	1	AF210249	AF210249 Streptomy
C 15	110.4	22.0	292200	1	SCO939129	AL939129 Streptomy
C 16	107.8	21.5	12730	1	AE005055	AE005055 Halobacte
C 17	101.2	20.2	300425	1	AP005022	AP005022 Streptomy
C 18	98.8	19.7	2520	6	AX172316	AX172316 Sequence
C 19	98.8	19.7	9984	1	ASP290449	AJ290449 Amycolato
C 20	96.2	19.2	299300	1	AP005026	AP005026 Streptomy
C 21	92.8	18.5	301675	1	AP005027	AP005027 Streptomy
C 22	91.8	18.3	1578	6	E22246	E22246 gene DNA en
C 23	90.8	18.1	13040	1	AE005769	AE005769 Caulobact
C 24	90.8	18.1	349746	1	EX572605	EX572605 Rhodopseu
C 25	89.4	17.8	309350	1	AP005215	AP005215 Corynebac
C 26	88	17.6	295150	1	SCO939125	AL939125 Streptomy
C 27	87.8	17.5	18876	6	AX803758	AX803758 Sequence
C 28	87.8	17.5	61944	6	AX803750	AX803750 Sequence
C 29	87.2	17.4	22449	1	SPSNECDE	Y11548 S. pristinae
C 30	87.2	17.4	22449	1	SPSNECGEN	X9690 S. pristinae
C 31	84.2	16.8	11776	1	AE004683	AE004683 Pseudomon
C 32	84	16.8	300400	1	AP005943	AP005943 Bradyrhiz
C 33	83.2	16.6	346259	1	EX640435	EX640435 Bordetell
C 34	82.2	16.4	1992	8	AF052222	AF052222 Lolium pe
C 35	82.2	16.4	2038	8	AF052223	AF052223 Lolium pe
C 36	81.6	16.3	299950	1	AP005963	AP005963 Bradyrhiz
C 37	81.6	16.3	346287	1	EX640450	EX640450 Bordetell
C 38	81.4	16.2	7788	6	AX803724	AX803724 Sequence
C 39	81.4	16.2	37360	6	AX803722	AX803722 Sequence
C 40	81.4	16.2	301482	1	AE016916	AE016916 Chromobac
C 41	80.8	16.1	1749	6	AR384076	AR384076 Sequence
C 42	80.6	16.1	1623	6	BD179585	BD179585 Highly th
C 43	80.6	16.1	2006	6	AX204868	AX204868 Sequence
C 44	80.4	16.0	11544	1	AF320253	AF320253 Azorarcus
C 45	80.2	16.0	1979	8	AK120964	AK120964 Oryza sat

ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AMM223012 53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.

AJ223012.1 GI:2764760  
ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.

Schupp, T., Toupet, C., Engel, N. and Goff, S.  
Cloning and sequence analysis of the putative rifamycin polyketide

synthase gene cluster from *Amycolatopsis mediterranei*

Unpublished  
2 (bases 1 to 53784)  
Schupp, T.  
Direct Submission  
Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
SWITZERLAND

FEATURES  
Location/Qualifiers  
1..53784  
/organism="Amycolatopsis mediterranei"  
/mol\_type="genomic DNA"  
/strain="LBG A3136"  
/db\_xref="taxon:33910"  
1336..15543  
/function="CoA ligase, acylcarrier protein,  
b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"  
/note="ORF1  
PKS modules 1-3"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAAL1035.1"  
/db\_xref="GI:2764761"  
/db\_xref="GOA:O54666"  
/db\_xref="SPTREMBL:O54666"  
/translation="WRTDLIKPLHVALLENATRFAGKPAFPADHRTVTYGDLEARTTR  
LAGHLGVRHGRBVAICLNGRSTVSEYFAILIRAGAGVPLNGSATAELEHPLTD  
SGATVVTDAAQAKRLAPHELVLLTGDDVPEGNASHDELSALPEAPADDLELDE  
PAMFYTSVTGTRPGKGVSTQNCNLSVASCYVFPGLSDQDRLWPLPFLHSLHIA  
CVLATVYGVASVRIADGSSADDVMRLIEASTFLAGYPTTYHHLVRAARGFSAPS  
LRI GLAGAVLGLAGLRSFEETFGVPLIDAYGSTETCGAITMNPDPDGRVSGCLAV  
PGDVRVDPDTGLDVPAGEGEVWVGNGVNLGYHNSPEATAMARDGWFRTGLAR  
RDDAGYFTICRIKELIIRGANNIHPGEVEALRTVCGVADAAGVPHDITLGEVPA  
VYIPEPTGPPAALIEKREQLSAYKVPDRILEVAHIPTASGKIRGLITDEPAVRA  
YAAETEHSQSDHESVAAALRAKRLSGUDERACQELLEMLVRTQAADVILGQVPDGRA  
FRDLGFTSLAIVLNRILTEHTGLWLPASAVDFDHPFPAALAAVRVAELIGITQAVAP  
VVAADPGEPILAIVGMACRLPGCVASPEDLWRLVRAERVDVASEFPDGRWDLDLSDPD  
REKATSVYGGGFLHDAGEFDAGFGFI SPRAVAMDPOORLLLETSWEALENAGVDP  
TALKGDTGVFSGLGGQYGSAGVAPELEGFTVTCGASVAGSVYVILGLEGPATV  
DTACSSSIVAMHLAAQLRQGCSCMALAGGVVMTATPGSFVEFSQRALAPDGRCAF  
AAADGDTGSEGVVWVLERVSVAREGHRILAVIRGSVNDQGNASNGTAPNGLSQQ  
RVTIRALAAAGLSDVDVVEAHGTGTGLDPIEQALLIATYQKRGKPLMIGLSKSN  
IGHAAAGVAGVIMQVQALRHETLPTLHVDPKPTLEVDMSAGAILLELTFEAWPRNG  
RERAGSVSGVSGTNAHLILEEAPAEVPAPELPVPLVVSARSSTLSQAERLA  
SLLEGVSLTEVAGALVSRAVLDERAVVAGSREAVTGLRALANTAGSTPGKVWV  
FPGQGTQWAGMRELLASPTFAERLAECAALAPWIDNSLVDLRGSDLGRVDVLO  
PACFAVMVGLAAVWESVGRVPDAVGHSGQETAAACVSGALSIEDAKKVALRSQAIA  
AELSGRGGNVALGEDDVVSLVDGVEVAAVNGSPSVIAGDAHALDAITLISGEG  
IRVRVAVDYASHTHRVEDIRDTLAETLAGISAQAPAVPFYVTSSEWRDAGVLDDG  
YVVRNLQVRFCAATALLBOGHTVFVEVSAPHTVQPLSELTGDAIGTLRRDGL  
RELLASMGELFVRGIDVMTAMPAGWVDLPTYAFEHRRHYLWLEPAEPASAGDPLIGT  
VYSTPGSDRLTAVQWSRAQWAVDGLVPNAALVEAAILRLGLAGTPVPGELVDAP  
VILPRGSEVOLLI VEGPEQRRRPIEVFSREADEPWRKHAGTLPAAAAVPEPAA  
GDATVTVAGLADRGDTHPALDAAVTVVGGDLPPLSVTGWSSLASGATVATVTP  
TATGLRLTPAGQVLTAVESVGTGPVFAQGTGTDALFRVDMPEIPLPTAETADFLPYE  
ATSATLTSALQAWLADPAETRLAVVTGCTPFGAAAILWGLVRSQSEHPGRILFADL  
DPAVLPAVASEGPQVRNGVAVSPRLTRVTPRODARPLDPEGLVLTGGTGLTGA  
LTARHLVTAGVHVLVWVRGGEAPBLQEELTALGASVALACDVAADRAQLAELRAL  
PABHPLTVAHTGVLVDDGVVTELTDPRLATVRRPKVDAAALDELTREADLAFLVIF  
SSAAGVILGPGQGYAAANAEALDARQNSLDLPVSLANGVWATVSGMTEHLGDAD  
LRNQRIGMSGLPADGEMALLDAIATGTLVAAKFDVAALATAKAGGVPPLLRGL  
APLPRAAATASLTERLAGTEQAAILDLVRRAAEVLEHSGAESVHSQRTFKO  
AGDSLTVELNRLAACTLSLSPAMIDYPPKPPALADHLRAKLGSAANPAEITGT  
AAAEPIAIAMACRFPFGVHSFEDLWRLVADGADAVTEFPADRGNDTBLIHEDPDH  
EGTYVYHGAFLDDAAGFDAAFFGIPNEALAMDPOORLLLETSWELFEFRAADPTTL  
ACQDYGAVGNSHDYSMRHRAAGVEGPEFLITGCSASVLSGRVAYHFGPEADPTVT  
ACSSILVALHMAQLQRGCSCMALAGVMMGTVEFVFSRQGLAPDGRCKAFAD  
GADGTGSEGVGLVLRLESEAQRHGOVLAVVRSVNDGASNGLTAPNGSQOQRV  
IRKALAAAGLSTSDVDAVEAHGTGTLIDPIEABALLATYQGNRETPMLGWSVKSNI  
HTQAAAGVAGVIMQVQALRHETLPTLHVDPKPTLEVDMSAGAILLELTFEAWPRNGH  
RRAGSVSGVSGTNAHLILEEAPAEVPAPELPVPLVVSARSSTLSQAERLA  
AFLGDRDTRVPRDAAVALATTRAQLDHRVAVLVSADRAQLCADLAAFSGSVYTGTPVDG

CDS  
15550..30759  
/function="b-ketoacyl-ACP-synthase, acyltransferase,  
dehydratase, ketoreductase, acylcarrier protein"  
/note="ORF2  
PKS modules 4-6"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAAL1036.1"  
/db\_xref="GI:2764762"  
/db\_xref="GOA:O52789"  
/db\_xref="SPTREMBL:O52789"  
/translation="MANQSWRKMSAPNEQIVDALRASLKENVRLQENSAALAAAAE  
PVAIVSMACRVAGGIRGPDPRVVSSEGADVTPGPEDRGMDVEGLYHPDPNPGTTY  
VREGLAQDQAFDAGFGFI SPREALAMDPOORLLLETSWELTLAGIDPHSVGRSDI  
GVYAGVQDYADPLSGEGEFNMSLERALGTAGVASGRVATYTLGLEGPATVDTWCCSS  
SLVAITHLAQALRRGCSCMALAGGVVMTATPGSFVEFSQRALAPDGRCKSAALADG  
SWAGSEGVVLLILERSVAREGHVLAIRGSVNDQGNASNGLTAPNGSQOQRVIRKA  
LASAGLTPSDVTVEGHGTGTVLIDPIEQALLIATYQGNRQDQDPLMIGLSVSVGHT  
QAASVAGVIMQVLSIRHQLPATQVDAPTPOVDWSAGAILLEAEGRPNRHPRR  
GGISLADVAGTILVSRGMLSEVFAVVAAGHBEAVTGLRALARGESAPGLISRGSGVP  
GVVSVFPQGTQWAGMRELLASPTFAERLAECAALDELTREADLAFLVIF  
KGVVVPQASFAVMVGLAAVWESVGRVPDAVGHSGQETAAACVSGALSIEDAKKVAL  
LSQAIALSAGRGGNVALGEDDVVSLVDGVEVAAVNGSPSVIAGDAHALDAITLISGEG  
IRVRVAVDYASHTHRVEDIRDTLAETLAGISAQAPAVPFYVTSSEWRDAGVLDDG  
YVVRNLQVRFCAATALLBOGHTVFVEVSAPHTVQPLSELTGDAIGTLRRDGL  
RELLASMGELFVRGIDVMTAMPAGWVDLPTYAFEHRRHYLWLEPAEPASAGDPLIGT  
VYSTPGSDRLTAVQWSRAQWAVDGLVPNAALVEAAILRLGLAGTPVPGELVDAP  
VILPRGSEVOLLI VEGPEQRRRPIEVFSREADEPWRKHAGTLPAAAAVPEPAA  
GDATVTVAGLADRGDTHPALDAAVTVVGGDLPPLSVTGWSSLASGATVATVTP  
TATGLRLTPAGQVLTAVESVGTGPVFAQGTGTDALFRVDMPEIPLPTAETADFLPYE  
ATSATLTSALQAWLADPAETRLAVVTGCTPFGAAAILWGLVRSQSEHPGRILFADL  
DPAVLPAVASEGPQVRNGVAVSPRLTRVTPRODARPLDPEGLVLTGGTGLTGA  
LTARHLVTAGVHVLVWVRGGEAPBLQEELTALGASVALACDVAADRAQLAELRAL  
PABHPLTVAHTGVLVDDGVVTELTDPRLATVRRPKVDAAALDELTREADLAFLVIF  
SSAAGVILGPGQGYAAANAEALDARQNSLDLPVSLANGVWATVSGMTEHLGDAD  
LRNQRIGMSGLPADGEMALLDAIATGTLVAAKFDVAALATAKAGGVPPLLRGL  
APLPRAAATASLTERLAGTEQAAILDLVRRAAEVLEHSGAESVHSQRTFKO  
AGDSLTVELNRLAACTLSLSPAMIDYPPKPPALADHLRAKLGSAANPAEITGT  
AAAEPIAIAMACRFPFGVHSFEDLWRLVADGADAVTEFPADRGNDTBLIHEDPDH  
EGTYVYHGAFLDDAAGFDAAFFGIPNEALAMDPOORLLLETSWELFEFRAADPTTL  
ACQDYGAVGNSHDYSMRHRAAGVEGPEFLITGCSASVLSGRVAYHFGPEADPTVT  
ACSSILVALHMAQLQRGCSCMALAGVMMGTVEFVFSRQGLAPDGRCKAFAD  
GADGTGSEGVGLVLRLESEAQRHGOVLAVVRSVNDGASNGLTAPNGSQOQRV  
IRKALAAAGLSTSDVDAVEAHGTGTLIDPIEABALLATYQGNRETPMLGWSVKSNI  
HTQAAAGVAGVIMQVQALRHETLPTLHVDPKPTLEVDMSAGAILLELTFEAWPRNGH  
RRAGSVSGVSGTNAHLILEEAPAEVPAPELPVPLVVSARSSTLSQAERLA  
AFLGDRDTRVPRDAAVALATTRAQLDHRVAVLVSADRAQLCADLAAFSGSVYTGTPVDG



[illegible]



ARWTVERRADLGEQASNDVDVYVQVEAPVETSAKVTHYTAGAGGSVAVQPSANVPN  
 PERGLYHHTGDCDADLADARRRYTEQGITLAMCVFYIAEHRSPFIATLDRLOF  
 QFDAARVAGVKLVIRFAYTASAGGEDKERVILAHIDQDLPLLRANADVAVVQAGFV  
 GAWGEWYTKFNAGVSGATQANRAVVDKLLTALPADRAVQLTRPKFKALGGG  
 STRIGHNDCEFLASDDTDFGYEDPAEYEPYLEARTLPMGGGETCAVNPPTDCPTAT  
 AELSRPHWLYATDYODDYLRAWSAOGCLAEVTKALGYRTLRDLSVFPTTTRGAALP  
 VRVNLNDGYAAVNRQTNVNLVNTTGTALTELPJTADPRTWLPGARALAQDAQIPP  
 SLATQYALLLELADPALPERPEYSIQTANTGLWDARTGLNDLKKQITVS"  
 1825..2142  
 /gene="asm36"  
 /product="unknown"  
 /protein\_id="AA054114.1"  
 /db\_xref="GI:21449382"  
 /translation="MYTASAVLCARACGAPRAGATARWRFCTACKQAAVRLRKA  
 GSASHEHEDQLREALLAVARATLTETPPNSAQATLVGLARRVVELAAARDAGV  
 SGG"  
 complement (2129..3307)  
 /gene="asm35"  
 complement (2129..3307)  
 /gene="asm35"  
 /codon\_start=1  
 /transl\_table=11  
 /product="bicyclomycin resistant protein"  
 /protein\_id="AA054113.1"  
 /db\_xref="GI:21449381"  
 /translation="MAYKQDSAVGWLPAIILVALTGTTIETIDMSVSPFPDIAREIG  
 VGGPAVOLVTNLFGLGICLACHGPLSDRGPRPVLGGSAVMLGALCALAPTIG  
 LLTPARVQIGASAAVAVFVITGDIYRGPAAALRMVGLVNAWSALMTCAPLLGGAI  
 NAAWGNLAVCATLALLVFLFELPETRTARTATARSVADYRLILGSRAYL  
 ASVTGELLVAFTFAASPLYTAAGFVLTGALHLLVVAASFAPVSAFGIIP  
 LLGGPAGAVRMLGAVVGLGLLGGDPVAVTGVSVALFCVGAFCVPPVFDMSV  
 FPFLAGASSATMGLRALLVTLTLLTALAGALSTGGFPAPAAWAGAVAAVLLSLSR  
 "  
 complement (3450..4064)  
 /gene="asm34"  
 complement (3450..4064)  
 /gene="asm34"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator"  
 /protein\_id="AA054112.1"  
 /db\_xref="GI:21449380"  
 /translation="MARNPERRAALLDAVEVLAGEGARGLTFRANDQAGVPAGTA  
 SNYFAGRAEILTAGERVYQRLLOQPLGETPDGPRDSRVTELMHALVDRVSAPTG  
 FLALLELEARPELEVLITARVADLRNABHHAASGLPGDGTVALLWLNWL  
 ILERLTPLDLTDHORDLVLTALVDRLLAGQPEAGSPFDESCK"  
 4151..4723  
 /gene="asm33"  
 4151..4723  
 /gene="asm33"  
 /codon\_start=1  
 /transl\_table=11  
 /product="dihydrofolate reductase"  
 /protein\_id="AA054111.1"  
 /db\_xref="GI:21449379"  
 /translation="MRKLVIYFAASIDGVIAGEGHEHDFYPTAEPMEVHELRAEYPTT  
 PTHVRLYQMLDPTNRRYDVTLMGRGVSQPALDAGVGSFYAHLRQYVVSRLPADP  
 AVLVRAQPVGLVRLKEDGDLVWLCGADLAGQLPEIDELVVKYTPVVGAGGTPL  
 FRAGFAPRAFPRVDVRAFDHGGVVTYRPA"  
 complement (4854..5057)  
 /gene="asm32"  
 complement (4854..5057)  
 /gene="asm32"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AA054110.1"  
 /db\_xref="GI:21449378"  
 /translation="MWITQNAIRIPARRSSGRSNARRCAVRWPRPQPSATRAPGPRS  
 "

SANADSRARSGRDPRPTSLNRSA"  
 5415..6461  
 /gene="asm31"  
 5415..6461  
 /gene="asm31"  
 /codon\_start=1  
 /transl\_table=11  
 /product="RNA polymerase sigma factor"  
 /protein\_id="AA054109.1"  
 /db\_xref="GI:21449377"  
 /translation="MMIGISADGRRRRPVRWCVTEAFWTETDEDIKLAATLERDAEA  
 ANADSTRAYLKAIGREVRLLTADQVELARRVEAGLYAEHLDDLPAAAGSRPVELR  
 RVVRDQGRADHLEARNLRVLSIAKRYTGKGMALDVIOEGLNLVRAVEKREATAEE  
 EKFSVTAWIRQAITRALADQSRTIRLPVHVVEVVTKLARIERDLHRRNGREATAE  
 VAVELDVTPKVKVELRQHARTPVSLDQITIGAEADTSGIDLIADENAVAPLEAVSSALL  
 SADLRVLETLTPREAGVVRMYGFFDPGLVHTLDEVGRVYGVVTRERIRQIESRTAKL  
 QHPARSGLEGYLD"  
 complement (6501..9518)  
 /gene="asm30"  
 complement (6501..9518)  
 /gene="asm30"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome P450"  
 /protein\_id="AA054108.1"  
 /db\_xref="GI:21449376"  
 /translation="MVATGTRIIPGPKPLPLVGNLLDVLTSDDTDVDFLDCHREHGG  
 IVALTPAGORQVFASSELVAFSSDPGKGAHPALEQVRDFAGDGLFTARGDEPNW  
 GKARLLMPAFGFTAMRDHPFMDLTAQMLVWRERPGDHRIDVADDMTRLTJDTTA  
 LCAFGARFYNDRAHPFVDVAVRSLVEAGERAERLPVQVPELVGNQRYRDDIATM  
 NRIADGVAAARALPAGEREDDILLERMLTADPVGTGERLSARNRVYQLATFLIAGHET  
 TSGLLSPAVHRLLAHPVLRKADAVDGLDVRPAFEDLARLDYLQVIRETLRLHP  
 TAPALADPEAPALGHAIGAGEPVLWMLPTLHRDPVAVWRDPDVFDEPFAPRMDI  
 IPACNMPFGHGACIGRPFAEQEATLVIALVLRFDLALADPDHRLTIKQITLKP  
 DSVRRAPRADRPAGATATVETVQHPVPAHRTGTHLVYVYSGSGSGEGLARTIAG  
 DGAARWATSPALDDADRALPAGSPVVTI VSSSNGAPPDNAAHFVRLIQDGPDLSC  
 DVLVLCGNLMSATYQRYVPTLIDEAAGARLRERGAITDARADFQDWRWYEP  
 LWPLLSACGVEGEIYQRFVSVESDAADGLDGLASAVVLENRELVRGPDGSKHLE  
 RLDPDGTSYRTGDIYSLVLPQNHPDLVRRVAVRAVRAVTVVSSAPTGLVPPVGRAL  
 RVDELTRCVDSLSPAGAGVVARLAERPCPPRAELAAATGATLLERLFPSCAVD  
 LALAEILLPAPRTRLYSISAAEQRAEVALTVSVGTGVTGSLVSRVPGDRAVAGIAS  
 PPESFRPADNTVPVLLIAGTGTPAPFGFLRAALGGEPGAPALLFGCGPELDDL  
 YAEFAALGDWLEVDVRAYSRHPDGEVHVQRLWRRDRVRVELVDAGARVILCGDATR  
 VGPAAVEVLGRIEFGAGWLDALRAGRYATDVF"  
 9664..10245  
 /gene="asm29"  
 9664..10245  
 /gene="asm29"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator"  
 /protein\_id="AA054107.1"  
 /db\_xref="GI:21449375"  
 /translation="MSARVRREQVVAALRVKRDGTAATTAICAEADMPHGAFHY  
 CFRSQEYLTALTTLGLNTSLDAWPAISDPLPAVDNVRTMLAAFWTAFQADPERELV  
 LGELVTLQAREPELPAWQHREYVAKVAHLDRLAEEAGFDVSDRALLAEVLVSV  
 LDGVASSWFSHRDDELAKASLDQPAVMPAALTR"  
 10393..10857  
 /gene="asm28"  
 10393..10857  
 /gene="asm28"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AA054106.1"  
 /db\_xref="GI:21449374"  
 /translation="MTDTTTRHTFPVQNAALAVGAVFLLVGIAGFVPGLTNTDGLF  
 AGHSDAALLGVFKVYLAHLNLVHLAFGAAGIAAGLSKAPRGARFLICGGAIYAVLWL  
 YGVVVDHSTANFVPLNAADNWLHLGLAVGMIGLGAALGRNAAHYRAARSDD"  
 complement (10912..11937)  
 /gene="asm27"  
 complement (10912..11937)  
 /gene="asm27"









```

RESULT 6
AF007101
LOCUS
DEFINITION
  32870 bp DNA linear BCT 13-JAN-1998
  Streptomyces hygroscopicus putative pteridine-dependent
  dioxigenase, PKS modules 1, 2, 3 and 4, and putative regulatory
  protein genes, complete cds and putative hydroxylase gene, partial
  cds.
ACCESSION
  AF007101
  AF007101.1 GI:2624946
VERSION
  1
KEYWORDS
  Streptomyces hygroscopicus
SOURCE
  Streptomyces hygroscopicus
  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
  Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
  1 (bases 1 to 32870)
  Ruan, X., Staasi, D., Lax, S.A. and Katz, L.
  A second type-I PKS gene cluster isolated from Streptomyces
  hygroscopicus ATCC 29253, a rapamycin-producing strain
  Gene 203 (1), 1-9 (1997)
JOURNAL
  98085969
MEDLINE
  9426000
PUBMED
  2 (bases 1 to 32870)
  Ruan, X. and Lax, S.
  Direct Submission
  Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott
  Park Road, Abbott Park, IL 60064, USA
  Location/Qualifiers
  1..32870
  /organism="Streptomyces hygroscopicus"
  /mol_type="genomic DNA"
  /strain="ATCC 29253"
  /db_xref="taxon:1912"
  complement(108..1130)
  /note="orf5; similar to RapK"
  /codon_start=1
  /transl_table=11
  /product="putative pteridine-dependent dioxigenase"
  /protein_id="AAC38061.1"
  /db_xref="GI:2624947"
  /translation="MNPSSVILNGLTSVFENGARVVPVGVGNILGVNVASVCEYPT
  LDGCVPLEINMAPTAEPAEAVVMTDAESPHGERDGTVAHDGEYFPCAGRPVPTGR
  YTEATRAAYVMFELLEEGYSSVFRMNEIGDINRDAEGMEYRPFCEGSAEAPQ
  CRLEDFPAAATGIGSGGGGIAFYLLACRSGGHVHIENPRVPAHHPKRYCPAPRF
  ARATYLPRAADGGGVGVSVGTAIVLGHETAHGDILVKQCRLENIELVLSGNLA
  AHGISAGHGLTALENIKYVVRSEDVPAVREICREAFSPDADIVYLTVDVCRSDLVE
  IEGVVM"
  1454..9625
  /note="orf1; loading domain"
  /codon_start=1
  /transl_table=11
  /product="PKS module 1"
  /protein_id="AAC38061.1"
  /db_xref="GI:2624948"
  /translation="MLREILRPPELLKAAHARGSOTAFADSRGVSVAQLEARTR
  RLAGLQLRLLPDRRAIYIGNCVETTESIALARASAIQVFINPHSSTAEALYLD
  DSAARVITDTRILPQVRRLQARPHLVTVVYDGEAGVAFDAAETEPQOPARDL
  GLDIAWLYSGTGRPKGLVSTQSCLSGVAAS SAGVLGLSETDRVLPVPLFSL
  AHIVLAVTAVGTARITEAPDAEELHTRAEFTPLAGVMTYHYLLDAARDGDL
  SLPLRVCLSAIATSTLRAEFQFTGVLDDCVGSTCTCGLITANWNGTQAGSC
  GLPVLGVTRIVDETHESVETGEGELIHWGSPSLVGVXNHPATEALRGWYRGT
  DLRGDDIGENYHTITGLSELIRSGENITFEVEQVLRVPGVTDAAVVGSAHPTLGE
  VFLVFLPAADGDFDEYFAACREHAYKVPPELYEIGSPIRTSGSKI KKHAAQSP
  ARLAVSGSGSFELREWTAPVPEAGVDPGQATSAHRAGSVTAGLEAFGSAVDC
  PRGLSDYLVARITGLTRSGKAVLTHGADHPQAQAVWGLVSMRADHPADAPLV
  DADGGEATLAVSILASQPRALVRGGVLPRLARAGSGQVRSRDPDGTVPVPGAD
  QEQRAELAGHLVAAYGIRHVRLLARTAVPDLVARLTELGSATVQTVRAPGRASLAE
  FQTAGPVTVLQVAPETLPTTLTGEVELGPLQDALDELTRAHPALFVSLAIRH
  NGPAHVTQAVNRRHAAGLPALASWGPREELTSRERIALDATTYDDVASLAIH
  ELAASVDTGAAGRLAPAAERGALLDLVLSVEYVITGGRPVHDQSGFRDIGNWTS
  SAVERLGRLAATGLALPATLAFDVPNDALAAAYGGRLFGDTSAGPDTEPTDPAAP
  SGEPALVGMACRLPGGVTSFBEELWELISSGGDGI SGFTDRGWDLEHLHYHPDPHPHG

```

```

TSTVREGGFTHDAGDFAALFGISPREALAMPQORLLLETTWEVERAGIDPSTLRG
SRTVFAGVMYHDIISRLRTVPPELEGYSSGGSGSVVGRSVTFEFGFAGVSDVTA
CSSLVALHLAVALQARTGECTAVAGVTVMASSPSYIEFSRQGLSPDCRCKSFAAD
ADGAASEGATLILVERLSAQRGLGHPVLGVGTAVNODGASGLTAPNGPSQORVI
RDALTNAQLTATDIDILEAHGTGTGLGDPPEAQAVALMYQDRLDPWLSLKNIGH
TQAAGVAGVMKAVLSMRHGVMPKSLHIDEFTEVDMAGTVELLAEARDWPGADERP
RRAGVSFGVGTNAHVILEQAPPAPEPTESAEPEFSEAGSPVGLDELSPGSPVP
WLSAONETALKOLARQLVTEGTGVOIAHSLITQSTTLPHRAVLVSTREEFAVA
LEQITRHTPGGKTVLVPFGQSGOWMGORLWETSPLYRDTTHEIDTALSEYIDWSV
TDILTSQGAPTWDRVDVIOPLVETTMALATLWQHGLTPDALIGHSGEIRAAAYTA
GAULTQDAARITLRSKALLPLTGHGIMISVLAPEKIQELIADPDHWNQKILAPITP
PNTTITGDTQATITLTLQOLSHRIMRWQPLGVDFAGSHIDQLHNLQKILAPITP
QKPTIPFYSTNTNTWITDLDADYWRNLRFQVLPQAINHLTNNGYTIETSPHP
TLTPTQETWPNPTTITLHNNNDTHALITGALGHATHGHPITWHTLIPPTKTPLP
TYPOHTRVWLNKPTVTDQGLGILPPLPVAPVTAETGILNHPINPAHH
PWTDHANGTPLPGTALYDMALHAGHDHHTLDELITHTPTLTHPTTIQLTISF
PDNTHLTVTRNSPDGPWTAHATGTLIQHHTPDPTPQTTWPTTDAEPIPLDGTQ
RLAATTGLEYPFTQGLDMLWRDSRLFADVLPNTDHTSDAGYGIHPALLDAAHL
VAVSDGDEIRLPFAMTGTILHATGKRLVELEWDSASGLAFDPGSGQPVVTVESL
ASRVISPEQLRSASARGPLYPRIEAWSGAGTPGAVERPEYEVPTGVDDVLADTR
RSTETLRIQOHLAADTTPLVAEATHDDLAGAAVMGLIRTAQTEHPRGILIDTNRH
PYSKALPTILIASGEPQARIDGAILPLRTVTVAPTEPAGVGECTVLTGTGSLGSL
IARHLVTEHGRDIVSVRQGPDRAPGATELIELQQLGAQVRVSCITTDRAALALV
DEVGPFVGVHTAGALADATVDMLDADALATTAAKAAHAWLHEHTQDGLTLMVF
SSLASVLSGAGQANYAANSFLDALATHRETLGLPGTSLAWSGWERTSGMSAHLTRAD
HDLARAGIRPLTDEQGLEPFDGALAVDLQGVVATFDLPAIGRSGAVPVLRLGLVRA
PRRAAGCAGSGAEPTERATLAPDARERSVLTSEVQAQALGHPDPGAIEPDQAF
KSLGFDLSLTAVERNRLSANTGLQLPATLVDFDPAAGATVLLDDVSPAPAAAPDD
DPDERELRVLRVFPVARFVREAGVIEALLKLAGEPDPARTDPAEDEDGLDAMTEA
LIRHVMDDTGA"
9637..14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEQLVEALRVSAKENARLRRENASIQEAGAEFPAIVGMACRL
EGVDSPESLWELMASGDVAPPTDRDMDGLXDPDPDHLGTAVYREGGFLDRAG
VFPDPPGIGSPREALAMPQORLLLETTWEVERAGIDPPTLRGSRGTGFMVQDY
VTRDLPFALEGYLNGSASVASGRVSYTFEGEFAVSLDTACSSILVALHLAVRP
ALRELLARAGVTVMSVPHYPVFSRQGLAFDAPCKFSQQRVIRTDALTDAELTAVD
ERISDLARLGHPLVAVRGTVAVNODGASGLTAPTSQQRVIRTDALTDAELTAVD
LLEAHGVTGLGDPTEAQUALIATYQDQDQVWGLGSLKSNIGHTQAAAGVAGVKTVM
AMEHGMWPTLHVSEPTNVDSAGEVRLITWARDSPSTDQORPRAAVSSFGISGTNA
HVLUQAPADPEPDPVTPPAGVPVWLSAHTGALSALAEALAAHVEEDRAPLGDA
HTLNTTATLPHRAVVGVEETREFFAALRAVRPRRGRPLAVMTGGSORPQMGRO
LYEHFPFAQALDEVFALATPLGRVMPDPAQETLQRTDHAQALFAFETALITLME
SWGLRPMVCGHSGEITAAHVSGTLTLPDAVHLVTRTGLTMQLNMLNVSHAFSLP
HTLOPHLNDHDTISIAAINGHATVLSGDRITLHATLQNTKNTNMLNVSHAFSLP
MOPILQPTTTLNLTTHHPHTPLISMLTATPHDPTHTHTOHTAPVRYDTLHLH
DHGTLVLEIGPDTLTLHARTITLPTTHLIPPTERNNEVRSNEALGRVFSVCHSV
DMRALTPTRGTSPLTPFFQRDFWLDAGGAEVGAAGLGTDDHPLGLAVTVADT
GELISGLSTSTHPTLTVNGTIVPGTALHSDRTTGTWTRHTTGLTSQTHPAELSWSPT
ALHTPSSQTSVGAETDTHRPIALHSRDTTGTWTRHTTGLTSQTHPAELSWSPT
DAHOLDITTAQOALDGLHGVPAQGLHTLYQRDNTLPAELTEPAGLPTGHLIHP
ALLDAAHPAASVARSARDESGIQLPFSWAGTVHSTGATHRLVRDLTDDTCAIRLAT
ALTEPAGLPTGHLIHPALHSTGATHRLVRDLTDDTCAIRLAT
DTTCQPVITISLTPRLHPOLHAAQEPETHLITLWSPPTLPTTTGHERVTVPT
AGDILADTHRTITATLRIQOHLAADTTPPLVTEATHDDHDLGLAGVWGLIRTAQTEHP
GRILIDTNDHPTSRDALASVAVSGKQARIDGAILTLPRLTRUSVDTRASTREHPL
SMGVLVTGGTGGGLSLIARHLVTEHGVLDLVLSRQGPADAGTALTNLQKQHGARI
RTISCULDRTALTTLDEIGLTVGVHTAGALADTTHLNDPDTLTTTLTKANPAW
WLHSLTDHDLALFVPSVAVGNAGQANYAANSFLDALITHRRQGLPGLTSLAW
GHWOREGTHLITQADHDMTRAGIHLTDAEGTTLPTDLDRLGLGAVAPVUHLHP
LNRADTPAVLRGLVRAVPAARQATVAESWADRILAAHMEERDRAVLDLVROQSR
VLGSHDADRIDPQGAQFQDFTGDSLTAVERNRTATGLRLPATLIFDHPITNALADH
LLAQLGPDGSGSPVVRDLERFERQLFTDRPDEVDGTAIVRLERSILAEHLADPAT
APAARTASQVLDITLEDASQSELLSFIDKEFGASN"
14964..21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"

```

CDS

CDS

CDS













Db 6410 CTGCTCTGGCCCGCCCGCTGTTCCACAGCTCGGGCACCATCTCTGCGCTCGCTCGCGGTC 6469  
QY 181 ACCGTGGTGGGCGGACAGCTCGGATCGCGACCGGACGTCGCCGACGACGTCGATCGCG 240  
Db 6470 CTCACCGTCCGCGCGCTCGCGCCGATTCCTCGCGGCTTCGTCGCGGCTGACGTCCTCGAC 6529  
QY 241 CTGATCGAGCGGAGAGCTCGACTTCTCTGCGCGGCTGCGGACCGACCTACCAACCACTG 300  
Db 6530 GCCCTGGCCGAACACCCCTGCACCGCTGCTGCGCGCTGCGGACGATGACCGTACCTC 6589  
QY 301 GTGCGGCGCCCGCGGACGCGGCTTCTCGCGCGGACCGCTGCGGATCGGCGCTGCGCGG 360  
Db 6590 CTCGTGCGG---TGTCGGGGAACCGCAGACCGCGCGCTGCGGCGGCTGCGGTCGCG 6646  
QY 361 GCGCGGCTCTCGCGCGCGGCTGCGGAGCGAGTTCGAAAGACCTTCGCGGCTCGCGGTCG 420  
Db 6647 GGGTCCACCTCACCGCGCTCGCTCACGCGGCTTCGAAAGCGGCTTCGAAAGCGGCTGCGGCTG 6706  
QY 421 ATGACCGCTACCGCAGCAGCGGACCTGCGGCGGATCACCATGACCCCGCGGACGCGC 480  
Db 6707 CTGACACCTACCGCTGACCGGACCGACCGATCGCTCACCGGCAACACCCCTGACCGAC 6766  
QY 481 GCCCGCTCGAGGCTCGTGC 501  
Db 6767 ACCAGGCTCCCGGCTCGTGC 6787

RESULT 11  
AF235504/c 77534 bp DNA linear BCT 19-JUL-2000  
LOCUS Streptomyces hygroscopicus var. ascomyceticus FK520 biosynthetic  
DEFINITION gene cluster, partial sequence.  
ACCESSION AF235504  
VERSION AF235504.1 GI:9280381  
KEYWORDS Streptomyces hygroscopicus subsp. ascomyceticus  
SOURCE Streptomyces hygroscopicus subsp. ascomyceticus  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
REFERENCE 1 (bases 1 to 77534)  
Wu, K., Chung, L., Revill, W.P., Katz, L. and Reeves, C.D.  
The FK520 gene cluster of *Streptomyces hygroscopicus* var.  
ascomyceticus (ATCC 14891) contains genes for biosynthesis of  
unusual polyketide extender units  
Gene 251 (1), 81-90 (2000)  
20323220  
10863099  
REFERENCE 2 (bases 1 to 77534)  
Reeves, C.D., Wu, K., Chung, L., Revill, P. and Katz, L.  
Direct Submission  
Submitted (16-FEB-2000) Kusan Biosciences Inc, 3832 Bay Center  
Place, Hayward, CA 94545, USA  
JOURNAL Location/Qualifiers  
TITLE 1. 77534  
FEATURES  
source  
/organism="Streptomyces hygroscopicus subsp.  
ascomyceticus"  
/mol\_type="genomic DNA"  
/variety="ascomyceticus"  
/db\_xref="taxon:132248"  
complement (412..1836)  
/gene="fkbw"  
complement (412..1836)  
/gene="fkbw"  
/codon\_start=1  
/transl\_table=11  
/product="fkbw"  
/protein\_id="AAF86380.1"  
/db\_xref="GI:9280382"  
/translation="WVLPIGVGRAPRRRTGALFLAAVLLPSPATARPAPDAPDG  
HCARLSQRVFAAARQQADCLDELTTAGTATGCTDPADWAGLTGPKNLATPTGPGVIG  
IDGYPFTSTTNHGMNHAQFVRLPDRMNGGLVAGTFCNREQVANDRAIDAWVL  
ARGYAAATDKNGTGPFAVRDGRPGDAVAEWNTRLTQLTRAARASVAQRVHRPLERT

LVTGMSGGLYVRWQLEHPELYDGGVDWEGTLMRAEGPNLFTFLPAALRDYPTYAAG  
GPAAGCARALLRAGPADSEFLVPYHVEVWDLTORVYREEDPFGDGEKEAGTPTYC  
ACTVPGCADVDYATRSSGVHAAVERIALTGRIKPLITIOALSTCCCLPPGLRRVR  
AYVRPGGRYRYRIEDSTHLDALVDTPDLRLPLLCHRSFTALESWLTGHRPP  
ANHVVPRGADQASULLACRLDG"  
complement (2020..3579)  
/gene="fkbw"  
complement (2020..3579)  
/gene="fkbw"  
/codon\_start=1  
/transl\_table=11  
/product="polyhydroxybutyrate depolymerase"  
/protein\_id="AAF86381.1"  
/db\_xref="GI:9280383"  
/translation="MOPPPRGLITPLPLSSPPVGSLSRPGRRGVLTRIVAVVAVL  
LGAALGPAPTAHAAGLAKGLTKADLTEVADFGTNGRLNMVYVRPASLPAEPVAV  
FAUHGCTQDQAGYADNSGLLSPADRYGFLVFAETSSNNANCFNWFQSSDNRGCG  
EASIRQMAHTVAYGADPQRTIITLSAGGAMTSMVLATYIDVFOGAVVAGLPFG  
CATDVSAYLCMNPFGDLDADQAWRRVDPSPSGWPRVAIWHGDKDTVAPRNAD  
ELRDQWAVHGVSTPDRTSVIGPNSRHEEYLADSGSAVAVNRPVGIHGTPVDPG  
TGAQCGSTGAAYELDSICSSYWTQFGLSGSADPSGLPAPSLAATGATDTTISL  
TWKPDGATYAVHRDQOITTSATSYTDTGLRAGTSHYVAARDADKAGPLSA  
VTAQTGTATVACWTAAGYAHVQAGRAITSAGYIYAKSGGQMGLYNTFTVTLKESPT  
GYFTVANDTCP"  
complement (3969..4496)  
/gene="fkbw2"  
complement (3969..4496)  
/gene="fkbw2"  
/note="transcription regulator"  
/codon\_start=1  
/transl\_table=11  
/product="fkbw2"  
/protein\_id="AAF86382.1"  
/db\_xref="GI:9280384"  
/translation="WVEGWQGRIFEDFTGDVYRHPGLGRVLTETDNSWLTLLTONTAP  
LHFDHYSGATGKPLVDFTFLALVTGQSVTDLSQHVMANLGWDRVLEPNFEGD  
TIYQSSEVLNTRSAASRPVGIISVRTIGYNQDGIIVITFERTLMVYRGGHGRFAAV  
EPWDESRRAAGLH"  
complement (4595..5488)  
/gene="fkbw1"  
complement (4595..5488)  
/gene="fkbw1"  
/note="transcription regulator"  
/codon\_start=1  
/transl\_table=11  
/product="fkbw1"  
/protein\_id="AAF86383.1"  
/db\_xref="GI:9280385"  
/translation="MEIRELQWFTALAEHEHVTITAEKLNISQPTLSAIRRLRRVVG  
VPLFDRHONRLNLYGVEFRAHALRAISEVGAQRITALIDPDTGTALGFLHSYG  
TMLVPDLLAGYRRLAPGTTPELRGDAADAVDDVNSRLDGLTSPRPGADVEWTPPL  
RDEALCLLVPFCHRLAPRRVRVTEALADEAEVALEPVGRLQITDRLCAAAGFVPRV  
LESTELSLRALVAAGLGVAVAPLITGSPSGGDHNAVPLDDPQARRVGVVTLAGGP  
RAPVVARFYDVRSRAHAPHV"  
5601..6818  
/gene="fkbw"  
5601..6818  
/gene="fkbw"  
/note="L-carnitine dehydratase"  
/codon\_start=1  
/transl\_table=11  
/product="fkbw"  
/protein\_id="AAF86384.1"  
/db\_xref="GI:9280386"  
/translation="MSEPHPRPQERAPGLSGLLVLSLEQAVAPAPATRIHLADIGAR  
VIKIERPGSLARGYDTRVGRMSHFVNLNRKESVQLDVRSPGNRHLHLVDRAD  
VLVQNLAPGAARLASATRSRGATEADHLRTYPATVPAADKADYDLVLQCEAGL  
VSTICTPTPSKVLGIADIACGMVAYSGILITALKRATGRGSLQLEWSMLEALGEGW  
GYAEVYTGCTAPARACASHTAPYGFPTTRDQGTINLGLONREWASFCGVVLQR  
PGLCDDPFGSNADRVARTELDALVSEVTCGLTGEELVAREASIAVARQTVREF  
SEHPLODRGNWAPDPSVGALEGLIPPTVTHGEHPRRLGRVPELGEHTESVLMAIA  
PHSADREAGAHAE"  
6808..8052

gene





```

FEATURES
  source
    Location/Qualifiers
      1..77536
        /organism="unknown"
        /mol_type="genomic DNA"

ORIGIN
  Query Match          36.8%; Score 184.2; DB 6; Length 77536;
  Best Local Similarity 61.9%; Pred. No. 2.5e-13;
  Matches 310; Conservative 0; Mismatches 188;
  Indels 3; Gaps 1;

```

QY	1	ATGTTCTACACGTCGGGCACGACCGGGCGGCCCAAGGGCGTCGTCTCCACGACGGCAAC	60
DB	46081	ATGCTCTACACCTCCGGCACCAACCGGCCGTCCCAAGGGCGTGGTTTCGCACACGGCGACG	46022
QY	61	TGCCTCTGCTCCGTCGGCTCTCTGCTACGTGCGGTTCCTCCCGGGGTTCTCGGACGACGACCGG	120

[illegible]

	181	45901	241	45841	301
Qy	ACCGTGGTCCGGGCCAGCGTCCGGATCCGGACCGCAGCTCCGCGCAGCAGTGTGATGGG	CTCAGCGTCCGCGCGTCCGCCCGCTATCCTGGCGGCTTCGTCCGCGCGTGACGTCTCTCGAC	CTGATCCAGGCGGAGAGCTCGACCTTCTTGGCCGCGGTGCCGACCACTTACCAACCACTTG	GCCCTGGCGCAACACTCCAGCACCGTGTCTGTCCGCGTGCCGACGATGTACCGTACC--	GTGCGGGCGCCCGGAGCGCGGTTCCTCCGCGCCGAGCCTTCGGATCGGCCTTGGCCGGG
Db					
Qy					
Db					
Qy					

D <sub>b</sub>	45783	-TCTCGGGCCGCTTCGCGGCAGCCGCGGGCTCGCGGCTCGCGGTGCGCGTGTGCGC	45722
Q <sub>y</sub>	361	GGCGCGTCTCTCGGCGCCCGGGCTGCGAAGCAAGTTTCEAAGACACTTCGGGTCCC CGTG	420
D <sub>b</sub>	45724	GGATCACCTCGCCGCGCTCGCTCCACAGGATTTTCEAAGGCACATTTCGGGTGCCCGTG	45655
Q <sub>y</sub>	421	ATCEAGCGCTTAGCGGACGACCAGACCTCGGGGGCATCACCATGAACCCGCGCAGCGC	480

[illegible]

AF210249/c	AF210249	77457 bp	DNA	linear	BCT 29-AUG-2000
LOCUS	Streptomyces verticillus bioamycin biosynthetic gene cluster,				
DEFINITION	complete cds.				
ACCESSION	AF210249 AF149091				
VERSION	AF210249.1 GI:9937210				
KEYWORDS	.				
SOURCE	Streptomyces verticillus				

ORGANISM      *Streptomyces verticillus*  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE      1 (bases 74421 to 74693)  
AUTHORS      Du, L. and Shen, B.  
TITLE      Identification and characterization of a type II peptidyl carrier

JOURNAL  
MEDLINE  
PUBMED  
AUTHORS  
TITLE  
REFERENCE  
JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

20389599  
10930733  
3 (bases 1 to 77457)  
Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.  
The biosynthetic gene cluster for the antitumor drug bleomycin from  
Streptomyces verticillus ATCC15003 supporting functional  
interactions between nonribosomal peptide synthetases and a  
polyketide synthase  
Chen. Biol. 7 (8), 623-642 (2000)  
20500448  
11048953  
4 (bases 1 to 77457)  
Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.  
Direct Submission  
Submitted (01-DEC-1999) Chemistry Department, University of  
California at Davis, One Shields Avenue, Davis, CA 95616, USA  
On Aug 29, 2000 this sequence version replaced gi:5328869.  
Location/Qualifiers  
1. .77457  
/organism="Streptomyces verticillus"  
/mol\_type="genomic DNA"  
/strain="ATCC15003"  
/db\_xref="ATCC:15003"  
/db\_xref="taxon:29309"  
complement (1199. .2185)  
/codon\_start=1  
/transl\_table=11  
/product="Syrp-like protein"  
/protein\_id="AAG02342.1"  
/db\_xref="GI:9937211"  
/translation="MTGSVTLTPLGGIIPRPRGEGITGAEDYDLGDLGADGDWVRHA  
GRLRLATDGLILHLGDTGDDGVDGVDGVPVGGDLPVYTERSTPRSVVKNYI  
TSTEPADQPIPMHNNSVAHWPSTLYFCHTAPDTGGATPIADRAVLIDLPAEVR  
RESQGVYTRFRADMGLSMQEAQFQEDRGDVERHCRAGHGOEFSDGDLTRHRP  
LATDPGTGAEVWFNQAHLEHSSLDPLDRQVILETYGNGELPRDLAFALGPIPDAD  
LATVRAATRAALPFRREGDMLVNDLRMAHGREFFTGERRVLVAMTSADS"  
complement (2182. .3918)  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthetase NRPS12"  
/protein\_id="AAG02343.1"  
/db\_xref="GI:9937212"  
/translation="MTLIRDAFLDQARTPDAAHVHGGTVMTVRELELRAGRMAT  
LAAGAGPQTLVAVRLPRGPPEVAALLAVLTGAGVPLADDDPPDRCHILDDCAA  
LILAHPSRDLRTLTDEALAPARPDAAVRAGDPAYVITYSGSSRGKGLVVEQGA  
LGAYLAQARVDGLSGRTVLHSSLSFDMAVTSLMGLVSGGAHVLDLKALASGTOP  
PAAARPSFLKVPRLHLLGLLPSCLPTGQLVIGREALGSGALGPRWAHPDVTY  
VNEYGTEATVCCATVPRGDVAPVPIGRIFAGTRLYLDADGEPVAVGGVGL  
HIAGDQALRGYLRPLTEERFVPDPFAADGSMYRTGLVRRPDPDGLYLGADGQ  
VKVSGYRIEPIETEAIVLGRHAGVRDCAVAVGEADARLVAVVVPDPSPGTAAPAR  
HAAALPPYVMPATVTVPELPLTPNGKLDRLALPGPDAGDGPGRTPAETLLICLL  
ARALGPIEDADAFILTSQGSTITALKLAVAGRRVIGIRLELITVLRITVRRILAAQP  
DAASPLAEGVE"  
complement (3915. .4733)  
/codon\_start=1  
/evidence="not experimental"  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG02344.1"  
/db\_xref="GI:9937213"  
/translation="MSKNAHWSRIKRTGDAGPVVLAVDFYGTGQREATFRHLCDLLTD  
PVMVAVPPAPDGDMSATGAGLRWMTGLDVLVAGRPVAVLGVCGAGVPASALA  
DALVERGHPRVVLPNPSAPVATLTDPRGLIAGMDLLTDGERAALLAETVAIRRA  
HAPDLVPVAAVLYREGCDLLCERLGVDSAGFAELAVLHSLAYLTAALDVPT  
ELWRGAVSLTSREHQGDFDVEHGFDAVARAILSSPQVVAULTALLREHESR"  
complement (4741. .6024)  
/codon\_start=1  
/transl\_table=11  
/product="putative transporter"  
/protein\_id="AAG02345.1"  
/db\_xref="GI:9937214"  
/translation="MTARVDATPTTYIAVLAVREARAPLIGLSCILARMSFAVLIALLIS  
VRDATGSFAVAGLTSGALSATLTFAPARALLIDRRGSRGLVRLTVPYLLGLAVLIT

LAEEAPTALLVAAAVAGVFAFPPLGPTMRVLMVILHGRQPLHTTAVLSDVTEVV  
FTVGLLAGLIAVAAPLASMITYWVLTIAAGTACFVLSAATAAPASGEADEDRPHR  
PMALPGMTIVLSFGVGLVGVGLVQLVLPFTADHAGSPGAGCILLISLISAGSAGGLA  
YGRIMASTPVRRVGLVGTFTLAVPLCLTASVPVAGAFALLVGLICLAPLFTTAYLL  
XNDIVTASGAPTANTWSTANNGGFAAGSAAAGVLLDSRGPTVTVTAAFAVAAATA  
VMTVIRKTLILGAGHPPEAAATPADRTAPAEAE"  
complement (6021. .6773)  
/codon\_start=1  
/transl\_table=11  
/product="thioesterase"  
/protein\_id="AAG02346.1"  
/db\_xref="GI:9937215"  
/translation="WDFLRTVNMWFSGGCDGRPRVRLCALPYAGGTAAVFKDWPAAL  
PFGVELLTAHLFGRRDFTTEPPATLEATEARLCEALPSPDLPTVVLHSGMAGLYE  
VAARLAARGRAPNLLIAAACPFPVHPDPAAGPVEAEALATRAERPWDITLDELML  
EAVLPALVADITAGDRVHRPRPLDLPLKVIVGADDDGTDRWTLTGWRACTARCEV  
VVLPGHYFLETDRAAVLTATVTLAEAEVGA"  
complement (6784. .7797)  
/codon\_start=1  
/transl\_table=11  
/product="clavaminate acid synthase-like protein"  
/protein\_id="AAG02347.1"  
/db\_xref="GI:9937216"  
/translation="MIDPLRTVALTGESAEVDDLRLTLADVPVDSVGLHRLTELA  
AQELRLIRAEITGMRLYDSPRALVVTGVDDEIRIGPTPAARPAAPDPDKRLLELL  
LIHALGEAFGWTQQRGLVHDLVPVGEETAQMGSSSETELLWHTEADAFHPLRCL  
YVGLICLRNHRQAAATVGVPLSLRTEDRAVLEPRYLIRPTDTSHTPAQNTATGRSA  
ERFAAIAEMDDAPERVAVLFGDPEPYLIRPAYMSPAPGDAARAVDTVALLIIDE  
LRHVLDAGSLLLVNDYQAVHGKPFAPAAAYDGRDRWLKVNITRLRSRSARRSAT  
LIV"  
complement (7859. .9214)  
/codon\_start=1  
/transl\_table=11  
/product="putative transporter"  
/protein\_id="AAG02348.1"  
/db\_xref="GI:9937217"  
/translation="WMSKSRDLRDLGGEDPVVAQESPDQAGTFCQDDGLNVFAAL  
AALREVEVRPLPHHAGIRLHVHSSQAAVALGPMQVASTTGVAGDGHQAGNRGQ  
VTGCVGREGVAGDDELGAEDRPHARQLRDLDDLLMMAARLADLVNLTQTVQGGDL  
RGQVSHDLGDDVLPQGRGLSLGSLQRRRGDVGTAHASVGQPGPGPFAAAESCRG  
LYARQQDQALARAIVGEGFORGEDAGORVAVDHPDPVGEVGVGGQVREVGSQL  
GHRDRREVPVAGGDDVGVAGVGLRLAPVSAHVADRAVGHVGLLAVGREQQQ  
QGGAGRVHCPVDVLVQVEDLTHGGDRLLVADLLRFQRTGLVDHRRPVMPLARY  
DPGPGPSLAHSPLLTENSIPSTRTPRCHLKKRPKTSQSAARPRRTGRLPLVS  
H"  
complement (9211. .17736)  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthetase NRPS11-10"  
/protein\_id="AAG02349.1"  
/db\_xref="GI:9937218"  
/translation="NASDALSAHLEVSDCRQGGITVINDGSELAQNVAAVRFERY  
AIAPBETA LHKGAATGYDELNRRBELTATRLADAGAPSTLVAVALPRDPDLVATLC  
ALLKLGACILPDGPIAGCLREIMADAPDVLVYTRAVAPATGGVPLFDADPT  
CSAVLPHSGTAGSETAYVLYPTTPDEKSENSVSTRMARIYLDLDDTAGIPARAELIR  
LVAPLSGGRLVDADTEPRPTEAPEDMVEDVVAQVQVAVLGVDRVDRVGFDFDL  
GKGLAAVQVVARLRLKLVGLFELRALFDATPEELARVRAEAGCGQVREAALEP  
VGRSEPLPLSFAQORLWFLDRMPORAFVTCDAFRVCGIDILGALRRALMLVGRHE  
TURTAFERDGVQYOLVGADGGAARVAAPTRVDLSLELPAEREAELVAAEART  
PFRDAGALLRVVARLADDDHVLVSTHIIYSDAMSGLVDELGLRECVTGDP  
ALPPPQVADFAVQWAMAGVQEEHLYWKAUDLADGAPSVLRLPMDHPRPAVQSER  
GETVGPALPDVLAALKEKLGREGGATLFMTLIGAFQVLLARHAGQEDIVVGVFAAGRT  
RTETPLGVFVNTPLRAICAPGLSFRDLQVREALGAFQDLPEALVEALAP  
EDIGHNPLVQVTFOLLGTPAARPDILGTEVERYPVQEAQSDFDLSDIKRADDDGYR  
GLINYCPDLFDRERMEVLYGHVLTLLGAAADGPRPIGELPLSDGAEARLALGAVTRE  
DRAVAGSPVPERPVARATADARAVTCGATLTFEELNDRVRLAQLALGAVTRE  
TPVAVLRSTDSVVAALAMVAGGVIVELDDPMDADRTAYLILDDTAASVITRDLPA  
LFGRLVDRPRFAADGLVPAPRIDPDQAAYITYTSGTGAPKGVVVRHSLNHLTAL  
QATFLGHDPYLAGADGVPPGDAKRLTTLTAPTFDASMEQLSWMGLHSELFIVPDEV  
RQPSALVRFVREHRIIDVITTSQLELIVSHGLLDGEWAPSMVMVGVGVSPLMRTL  
RQRPFRCNLYGPTAETVATDCHDLSDDADVPVIGTLPHTHVRLVDRLRPVPGV  
AGEIYUGGLGARGYLNRLNRLATARRVADVPDPTFGSRLYRTGDRARWPDGTYLIG  
RTDDQIKIRGFRVPEGEIEAVLTHHPVKEAAVDDDAHARLVAVITLAEGGGAGPTDV





```
Db 268708 GCGGGGCGGGCGGACCTTCCTGCTCCCGGCTTCGACGGGGCGCGGATCGAGC 268767
QY 242 TGATCGAGCGGAGAGCTCGACCTTCCTGCGCGGGGTGCGGACCACTACCACTGG 301
Db 268768 TGATGGTGAAGGAGGGCGGTCAACACCTTCACGGGGTGCAGACCATGTTGTCGCCCTCG 268827
QY 302 TGGGGGCGCGCGCGGACGCGGGTTTCTCGCGCGGAGCCTGCGGATCGGCCTGGCCGGGG 361
Db 268828 ---CCGCCGCGGGCGCGGGCGCGCCCTGCCGAGCTGCGGGTGTGCATCTCCGGCG 268884
QY 362 GCGCGGTCTCTCGGCGCGGGCTCGCAAGCGAGTTCAAGAGACCTTCGGGGTCCCGCTGA 421
Db 268885 GTGCTCTGCTGCGGGTGGCGTCTCTCGAACGGTTCGAGAGGCTTCGGGGCGAATCT 268944
QY 422 TCGACGCCCTACGGCAGCACCGAGACCTTCGGGGCGGATCAACATGAACCGCGCGGACGGCG 481
Db 268945 ACGAGGGCTACGGGCTGTGCGAGACCTCTCCGGCGGCGGTGTGAACCAAGCCGGTGTTCG 269004
QY 482 CCGGGTTCGAGGGCTC 497
Db 269005 GCGCCAGGCGCGGCAC 269020
```

Search completed: June 17, 2004, 01:40:37  
Job time : 2283.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:08:13 ; Search time 285 Seconds

(without alignments)

7467.885 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 agttctacagtcgggac.....ccgggtcgagggctgctgc 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseq 29Jan04:\*

2: Geneseqn1980s:\*

3: Geneseqn1990s:\*

4: Geneseqn2000s:\*

5: Geneseqn2001as:\*

6: Geneseqn2001bs:\*

7: Geneseqn2002s:\*

8: Geneseqn2003as:\*

9: Geneseqn2003bs:\*

10: Geneseqn2003cs:\*

11: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	53789	AAV21187	AAV21187 Amycolato
2	302.6	60.4	14055	AAU61170	AAU61170 Actinosyn
3	302.6	60.4	82746	AAU61224	AAU61224 Actinosyn
4	243.8	48.7	2614	AAQ06631	AAQ06631 Gene conf
5	184.2	36.8	77536	AAU14651	AAU14651 Nucleotid
6	163.8	32.7	18660	AAU58472	AAU58472 Nucleotid
7	98.8	19.7	2520	AAH21138	AAH21138 Amycolato
8	91.8	18.3	1578	AAU25763	AAU25763 B.diminut
9	87.8	17.5	1584	ACA23280	ACA23280 Prokaryot
10	85.4	17.0	2006	ADC68433	ADC68433 Lolium pe
11	85.4	17.0	2014	ADC68538	ADC68538 Lolium pe
12	85.2	17.0	1530	ABK13572	ABK13572 Ryegrass
13	82.2	16.4	1992	ABK13573	ABK13573 Ryegrass
14	82.2	16.4	2038	ADC68539	ADC68539 S. arundi
15	82.2	16.4	2073	AAH42474	AAH42474 cDNA sequ
16	80.6	16.1	2006	AAU25197	AAU25197 Maize 4-c
17	80.6	16.1	2122	ADC68535	ADC68535 Lolium pe
18	78.8	15.7	1855	ADC68535	ADC68535 Lolium pe
19	78.4	15.6	110000	AAI99682_39	Continuation (40 o
20	78.4	15.6	110000	AAI99682_40	Continuation (41 o
21	78.4	15.6	110000	AAI99682_39	Continuation (40 o
22	77.4	15.4	2039	ADC68432	ADC68432 Lolium pe
23	77.4	15.4	2041	ADC68536	ADC68536 Lolium pe

24	77.2	15.4	2284	6	ABK13571	Abk13571 Ryegrass
25	76.6	15.3	30001	2	AAU61016	AAU61016 Total DNA
26	76.6	15.3	30001	2	AAU05110	Aax05110 S. aureof
27	76	15.2	7673	2	AAV58229	Aav58229 Omega-Cyc
28	74	14.8	1386	6	ABK73860	Abk73860 Bacillus
29	74	14.8	90600	6	ABQ78872	Abq78872 S. roseos
30	72.6	14.5	1934	9	ADC68537	Adc68537 S. arundi
31	71.6	14.3	135638	7	ABX34289	Abx34289 S. atroool
32	71.4	14.3	1986	4	AAH42473	Aah42473 cDNA sequ
33	71	14.2	1970	2	AAU25198	Aax25198 Maize 4-c
34	70.6	14.1	47988	9	ADE86070	Ade86070 Streptomy
35	70.4	14.1	7191	3	AAU97538	Aaa97538 Streptomy
36	68.8	13.7	88421	6	AAU40781	Aah40781 8842int 9
37	67	13.4	1704	5	AAH65415	Aah65415 C glutami
38	67	13.4	1827	4	AAU67757	Aaf67757 Corynebac
39	67	13.4	34980	5	AAH68525	Aah68525 C glutami
40	66	13.2	2124	3	AAU55810	Aac55810 S. lavend
41	66	13.2	2124	9	ADE10229	Adel0229 S. lavend
42	66	13.2	53500	3	AAU55842	Aac55842 Complete
43	66	13.2	53500	9	ADE10261	Adel0261 S. lavend
44	66	13.2	110000	4	AAI99682_34	Continuation (35 o
45	66	13.2	110000	4	AAI99683_34	Continuation (35 o

## ALIGNMENTS

### RESULT 1

AAV21187

ID AAV21187 standard; DNA; 53789 BP.

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX AAV21187;

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
XX P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX Claim 4; Page 53-102; 205pp; English.  
XX  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polypeptides, which can be used for assembling a  
XX library of polypeptides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 5.2e-77;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTTCTTACACGTCGGGACACGCGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 60  
DB 1825 ATGTTCTTACACGTCGGGACACGCGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 1884  
QY 61 TGCCTCTGTGTCGTCGCTTCTGCTACGTGCGCTTCCCGGGTGTTCGGACCGACCG 120  
DB 1885 TGCCTCTGTGTCGTCGCTTCTGCTACGTGCGCTTCCCGGGTGTTCGGACCGACCG 1944  
QY 121 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCGACATCGCTCGCTGCTGTCGCC 180  
DB 1945 GTGCTCTGGCGGCTCCGCTGTTCCACAGCCTTTCGACATCGCTCGCTGCTGTCGCC 2004  
QY 181 ACCGTGTTCGGGCGCAGCGTCGGATCCGACGCGCAGCTTCGCCGACGAGTCGCGG 240  
DB 2005 ACCGTGTTCGGGCGCAGCGTCGGATCCGACGCGCAGCTTCGCCGACGAGTCGCGG 2064  
QY 241 CTGATCGAGGCGGAGAGCTCGACTTCTGCGCGGCTGCGGACCACTTACCACTG 300  
DB 2065 CTGATCGAGGCGGAGAGCTCGACTTCTGCGCGGCTGCGGACCACTTACCACTG 2124  
QY 301 GTGCGGCGCGCGGCGCGGCTTCTCGCGCGGCGGCTGCGGATCGGCTCGGCTGCGG 360  
DB 2125 GTGCGGCGCGCGGCGCGGCTTCTCGCGCGGCGGCTGCGGATCGGCTCGGCTGCGG 2184  
QY 361 GCGCGGCTCTCGGCGCGGCTGCGAGAGCGAGTTCGAAGAGACCTTCGGGCTCCGCTG 420  
DB 2185 GCGCGGCTCTCGGCGCGGCTGCGAGAGCGAGTTCGAAGAGACCTTCGGGCTCCGCTG 2244  
QY 421 ATCGAGCGCTTACGGGAGACCGAGACCTTCGGGCGGATTCACATGAAACCGCGGAC 480  
DB 2245 ATCGAGCGCTTACGGGAGACCGAGACCTTCGGGCGGATTCACATGAAACCGCGGAC 2304  
QY 481 GCCCGCGTCGAGGGCTCGTGC 501

DB 2305 GCCCGCGTCGAGGGCTCGTGC 2325  
RESULT 2  
ID AAL61170 standard; DNA; 14055 BP.  
XX AAL61170;  
XX 22-SEP-2003 (first entry)  
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #1.  
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX gene; ds.  
XX Actinosynnema pretiosum.  
XX WO2003045312-A2.  
XX 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Disclosure; Page 45-53; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
XX Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;  
Query Match 60.4%; Score 302.6; DB 7; Length 14055;  
Best Local Similarity 75.2%; Pred. No. 3.5e-43;  
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 1 ATGTTCTTACACGTCGGGACACGCGGCGGCCCAAGGGCGTCGTGTCCACGACGCGCAAC 60  
DB 484 ATGCTCTTACACCTTCGGGACACCGCGGCGGCCCAAGGGCGTCGTGTCCACGCGCAAC 543  
QY 61 TGCCTCTGTGTCGTCGCTTCTGCTACGTGCGCTTCCCGGGTGTTCGGACCGACCGG 120  
DB 544 TGCTGTGTGTCGGTGCGGCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603  
QY 121 GTGCTCTGGCGGCTCCGCTGTTCACAGCCTTTCGACATCGCTCGCTGCTGCTGCTGCTG 180  
DB 604 GTGCTGTGGCGGCTGCGCTGTTCCACAGCCTGTGACATCGCGGGGTGCTGCGCCACC 663  
QY 181 ACCGTGTTCGGGCGCAGCGTCGGATCCGCGGCGGCGGCTCCGCGGCGGCTGATCGG 240  
DB 664 ACCCGCTGCGGCGGCGGCTCCACCGGCTGTCGACGGGCGCTCCCGGAGAGCTGCTGCC 723  
QY 241 CTGATCGAGGCGGAGAGCTCGACTTCTGCGCGGCGGCTCCCGGAGAGCTGCTGCC 300  
DB 724 GCGCTGACGAGGAGCGCTCCACGCTGCTCGGCGGCTGCGGCGGCTGCGGCGGCTG 783  
QY 301 GTGCGGCGCGCGGCGGCGGCTTCTGCGCGGCGGCGGCTGCGGATCGGCTGCGCGG 360



Db 784 GTGCGCGCGCGGTGAGCGCGGTTTCTCGCGCCGACCTGCGGGTGGTGGC 843  
Qy 361 GCGCGGTCTCGCGCGCGGTGCGAGCGAGTTGAGAGACCTTCGGGTCGGCTG 420  
Db 844 GCGCGGATCACACGCGCGAGCTGCGGCGCGCCGTGAGCAGCGTTCGGCGTCCGCTC 903  
Qy 421 ATCGAGCGCTTACCGCGACACCGAGACCTGCGGCGCGATCACATGAACCCGCGGACGGC 480  
Db 904 ATCGAGCGCTTACCGCGACACCGAGAGCTGCGGTCGATCACCGTGAATGCGCGACGGG 963  
Qy 481 GCGCGGTGAGGGTCTGTC 501  
Db 964 CCAGGTCGAGGGCTCCAGC 984

RESULT 3  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX AC  
XX AAL61224;  
DT 22-SEP-2003 (first entry)  
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX Actinosynnema pretiosum.  
OS WO2003045312-A2.  
XX PN  
XX PD 05-JUN-2003.  
XX PF 21-NOV-2002; 2002WO-US037547.  
XX PR 21-NOV-2001; 2001US-0332158P.  
XX PA (UNIW ) UNIV WASHINGTON.  
XX PI Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Claim 7; Page 105-152; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
CC ansamitocin ansamitocin biosynthetic gene cluster I  
XX  
SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;  
Query Match 60.4%; Score 302.6; DB 7; Length 82746;  
Best Local Similarity 75.2%; Pred. No. 2.9e-43;  
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
Qy 1 ATGTTCTACACGTGCGGACAGCGCGCGGCCAAGGCGGTGTCACGCGCGCAAC 60  
Db 36558 ATGCTCTACACCTTCGGACACACCGCGCGGCCCCAGGCGGTGTCCTCCACCGCGCAAC 36499  
Qy 61 TGCTCTGGTCCGTCTCTGCTAGTGCCTGTCCTCCCGGTTGTGCGACCGACCGG 120  
Db 36498 TGCTCTGGTCCGTGCGCGCTGCTAGCGCGCGGTGTCCTGTCGACCCCGGACCGG 36439  
Qy 121 GTGCTTGGCGCGTCCCGCTGTTCACAGCGCTTTCGACATCGCTCGCTGTCGCGC 180  
Db 36438 GTGCTTGGCGCGTCCCGCTGTTCACAGCGCTTTCGACATCGCTCGCGGTGCTGGCCACC 36379

Qy 181 ACCGTGTGCGGGCCAGCGTCCGATCGCGACGCGAGCTCCCGCCGACGACGTGATCGG 240  
Db 36378 ACCGCGTTCGGCGCCACACCCCGGTGTCGAGCGGCGCTCCCGCAGGAGTGTGCGCC 36319  
Qy 241 CTGATCGAGCGGAGAGCTGACCTTCCTGGCCGCGGTGCGGACCACTACCAACCTG 300  
Db 36318 GCCTCTGACGAGGAGCGCTCCACGGTGTCTCGCGGGGTGCGCAGCAGCTTACCAACCTG 36259  
Qy 301 GTGCGGCGCGCCCGCGAGCGCGGTTCCTCGCGCCGAGCGCTGCGGATCGGCTGGCGGG 360  
Db 36258 GTGCGGCGCGCGGTGAGCGCGGTTCCTCGCGCCGACCTGCGGCTGGGTGGC 36199  
Qy 361 GCGCGGTCTCTCGCGCGCGCGCTGCGAAGCGAGTTCGAAAGAGACCTTCGGGGTCCCGCTG 420  
Db 36198 GCGCGGATCACACGCGCGAGCTGCGGCGCGCTCGAGCAGCGTTCGGCGTCCGCTC 36139  
Qy 421 ATCGAGCGCTTACCGCGACCGAGACCTGCGGCGGATCACATGAACCCCGCGGACGGC 480  
Db 36138 ATCGAGCGCTTACCGCGACCGAGAGCTGCGGCTCGATCACCGTGAACCTGGCGGACCGGG 36079  
Qy 481 GCGCGGTGAGGGTCTGTC 501  
Db 36078 CCAGGTCGAGGGCTCCAGC 36058

RESULT 4  
AAQ06631  
ID AAQ06631 standard; DNA; 2614 BP.  
XX AC  
XX AAQ06631;  
DT 24-OCT-2003 (revised)  
DT 27-AUG-2003 (revised)  
DT 25-FEB-1991 (first entry)  
XX Gene conferring teicoplanin resistance.  
DE Dalbaheptide antibiotic resistance; probe; ss.  
XX Dalbaheptide antibiotic resistance; probe; ss.  
XX Actinoplanes teichomyceticus; strain ATCC 31121.  
XX EP39328-A.  
XX 28-NOV-1990.  
XX PF 14-MAY-1990; 90EP-00109074.  
XX 23-MAY-1989; 89EP-00109273.  
XX (LEPE ) GRUPPO LEPETIT SPA.  
XX Denaro M, Lorenzetti R, Moroni MA, Sosio MA;  
XX WPI; 1990-356009/48.  
XX DNA sequence conferring specified antibiotic resistance - obtd. as DNA  
PT fragment by Bam HI digestion of Actinoplanes teichomyceticus strain.  
XX Claim 6; Page 11; 16pp; English.  
XX The total genomic DNA of A. teichomyceticus ATCC 31121 was digested and  
CC ligated with pJ3702 (from S. lividans). The ligation mixt. was used to  
CC transform S. lividans 66 protoplast. One clone was selected for its  
CC ability to grow in the presence of 20 mg/l teicoplanin. The plasmid in  
CC this clone was designated pTR168 and was found to contain the sequence  
CC below. The teico-R conferring sequence can be used as a marker in DNA  
CC manipulations. It can also be introduced into dalbaheptide antibiotic  
CC producing strains to improve anti-biotic prodn. yields. It can also be  
CC used as a probe to detect corresp. regions in the genomes of other  
CC producers of dalbaheptide antibiotics. (Updated on 27-AUG-2003 to correct  
CC OS field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 2614 BP; 310 A; 984 C; 994 G; 326 T; 0 U; 0 Other;







```

RBS      978. .981
FT        /tag= c
FT        987. .2462
FT        /tag= d
FT        /product= "fcs"
XX
XX DE19960106-A1.
XX 21-JUN-2001.
XX
XX 14-DEC-1999; 99DE-01060106.
XX
XX 14-DEC-1999; 99DE-01060106.
XX
XX (HAAR ) HAARMANN & REIMER GMBH.
XX
XX Rabenhorst J, Steinbuechel A, Priefert H, Achterholt S;
XX PI
XX WPI; 2001-399437/43.
XX DR P-PSDB; AAB86276, AAB86277.
XX
XX New enzymes from Amycolatopsis, useful for synthesis of vanillin from
XX ferulic acid, also nucleic acid encoding them and recombinant host cells.
XX
XX Claim 8; Page 9-14; 18pp; German.
XX
XX This invention describes novel enzymes (I) from Amycolatopsis sp HR167
XX involved in the synthesis of vanillin from ferulic acid (FA). This
XX sequence represents a cDNA fragment encoding the enoyl-CoA
XX hydratase/aldolase (ech) and the feruloyl CoA-synthetase (fcs) which are
XX described in the method of the invention
XX
XX Sequence 2520 BP; 355 A; 895 C; 929 G; 341 T; 0 U; 0 Other;
XX
Query Match      19.7%; Score 98.8; DB 4; Length 2520;
Best Local Similarity 54.5%; Pred. No. 2.1e-08;
Matches 243; Conservative 0; Mismatches 197; Indels 6; Gaps 2;
XX
XX 2 TGTCTTCTACAGTCGGGCAAGACCGGGGGCCCAAGGGCGTGTCTCCAGCAGCGCAACT 61
Db 1438 TGATGTACACTCGGGCAGCAGCGGACGCCCAAGGGCGGATGTCAACCCAGCGCAAC 1497
XX
XX 62 GCCTTGTGTGCTGCTTCTGCTACGTCGCTTCCCGGGTGTTCGACACAGACCGGG 121
Db 1498 TCACCTGGAATGCGTCAACGTCTCTGTGTGAGA---CCGACCTGGCAGCAGCGGG 1554
XX
XX 122 TGCTTGGCCGCTCCCGCTGTTCACAGCCCTTTGCAATCGCCTGCTCTCTGTCGCCA 181
Db 1555 CACTGGTTCGCGCGCGCTGTTCACGCGCGCGCTCGGCATGCTGTGCTGCCACCC 1614
XX
XX 182 CCGTGTGTCGGGGCAGAGTCGCGATCGCGACGCGGAGCTCGCGCCGACAGCTGATGCGGC 241
Db 1615 TGCTCAAGGGGGCAGCGGTGATCTCTGCACTCGCGCTTCGACCCCGCGCGCTGCTCG 1674
XX
XX 242 TGATCGAGCGGAGCTCGACCTTCTGCGCGGCTGCGACCACTTACCACTGCG 301
Db 1675 CCGTGTGAAACAGAGCGGCTCAAGCTGCTGCTGCGGTGCGCAGATGACAGGCGATCG 1734
XX
XX 302 TCGCGGCGCGCGGCGAGCGGCTTCTCGCGCGGAGCTGCGGATCGGCGCTGCGCGGG 361
Db 1735 CGCGCACCGCGGCTGTGCGCAGCGCGGCTTCTCGAGCTCGGACCTGCTGTGCGGG 1794
XX
XX 362 GCGCGGTCTCTCGCGCGCGGGCTGCGAAGAGAGTTCGAAGAGACCTTCGCGGGTCCGCTGA 421
Db 1795 GCGCGCGGTGCGCGGACCTCGCGAGCGGCTACCTCGA---CGCGGGCTCGGTTG 1851
XX
XX 422 TCGACGCTTACGGCAGCAGCGGACCC 447
Db 1852 TGCAGGGCTACGGCAGTACCGAGGCC 1877

```

```

ID AAX25763 standard; DNA; 1578 BP.
XX
XX AC AAX25763;
XX
XX 08-JUN-1999 (first entry)
XX
XX B.diminuta pimelyl CoA synthase coding sequence.
XX
XX Pimelyl CoA synthase; pimelyl CoA; pimelic acid; coenzyme A; ds.
XX
XX Brevundimonas diminuta.
XX
XX JP11046763-A.
XX
XX 23-FEB-1999.
XX
XX 31-JUL-1997; 97JP-00205791.
XX
XX 31-JUL-1997; 97JP-00205791.
XX
XX (MITU ) MITSUBISHI CHEM CORP.
XX
XX WPI; 1999-208106/19.
XX DR P-PSDB; AAW99453.
XX
XX New pimelyl CoA synthase gene - useful for preparation of pimelyl CoA
XX from pimelic acid and coenzyme A.
XX
XX Claim 2; Page 8-10; 11pp; Japanese.
XX
XX This sequence represents the coding region for the pimelyl CoA synthase
XX protein from Brevundimonas diminuta strain IF03149. The sequence can be
XX used for the preparation of pimelyl CoA from pimelic acid and coenzyme A
XX
XX Sequence 1578 BP; 273 A; 502 C; 533 G; 270 T; 0 U; 0 Other;
XX
Query Match      18.3%; Score 91.8; DB 2; Length 1578;
Best Local Similarity 51.2%; Pred. No. 3.4e-07;
Matches 241; Conservative 0; Mismatches 227; Indels 3; Gaps 1;
XX
XX 1 ATGTTCTACAGTCGGGCAAGACCGGGGGCCCAAGGGCGTGTCTCCAGCAGCGCAAC 60
Db 535 ATCTGTGTTCACACGCGGTACGACCGGAACCCGAGGGCGGGTGGGTAGCCACTGTGCG 594
XX
XX 61 TGCCTCTGTCGCTCGCTTCTCTGTCTAGTGTGCTTCCCGGGTGTTCGACACAGACCGG 120
Db 595 ATTGTGAGCAGGTGTCATGCGGGTCAATTGACGGGGTTCAAGTGTGGAACACGACGCTG 654
XX
XX 121 GTGCTGTGGCGCTCCCGCTGTTCACAGCCTTTCGACATGCTGCTGCTCTGTCCGCC 180
Db 655 ACCTTGGCGGGCTGCCGCTGTTCATACGTTTCGCGCCAGACTTCGGTCTGAACATGGCG 714
XX
XX 181 ACCGTGTTCGGGGCAGCGTCGGATCGCGGACGTCGCCGCGAGCAGTCTGATGCGG 240
Db 715 TTCGCGCGCGGCGGCGCTCATCTCTGATGCCACGATTTGACGGGACCAAGCTCTCGAA 774
XX
XX 241 CTGATCGAGGCGGAGAGCTCGACCTTCTTGGCGCGCTGCCGACCACTTACCACTCG 300
Db 775 CTGATGTTGAACACCAAGCAACGGTGTTCGCGGGTCCGACCATGTACGCGCTC 834
XX
XX 301 GTGCGGGCGCGCGGCGGCTTTCGCGCGCGGCTGCGGATCGGATCGGCTTGGCGGG 360
Db 835 GTGCGGGCGCGCGG---CCGACACGACGCGCGCGCGCTGCGGTCTCGGAATTTCTGCG 891
XX
XX 361 GCGCGGCTCTCGGCGCGCGGCTGCGAGCGAGTTTCGAGAGACCTTGGGGTCCCGCTG 420
Db 892 GGTGCGCGCTTCCGCTTCTGAGGGCTTTGAGAGGCGTTCGCGGTGCCCCG 951
XX
XX 421 ATCGAGCGCTACGCGACGACCGAGACCTTCCGCGGGCGATCACCATGAACCG 471
Db 952 CACGAGGGCTACGGGCTTACCGAGACAGCCCGGACCGGTACCGTCAATACG 1002

```



DR P-PSDB; ADC68468.  
 XX New polynucleotide encoding polypeptides from *Lolium perenne* or *Festuca*  
 PT arundinacea, useful for modulating the biosynthesis of lignin, fructan or  
 PT tannin in a plant.  
 XX  
 XX Claim 1; SEQ ID NO 143; 240pp; English.  
 XX  
 XX The present invention describes isolated polynucleotides (I) encoding  
 CC proteins (II) from *Lolium perenne* and *Festuca arundinacea* which are  
 CC active in lignin, fructan and tannin biosynthetic pathways. Also  
 CC described: (1) an isolated oligonucleotide probe or primer comprising at  
 CC least 10 contiguous residues complementary to 10 contiguous residues of  
 CC (1); (2) a kit comprising the oligonucleotide probe or primer; (3) a  
 CC genetic construct comprising (1); (4) a transgenic plant cell comprising  
 CC the genetic construct of (3); (5) a plant or its seed, fruit or progeny  
 CC comprising the transgenic plant cell of (4); (6) modulating one or more  
 CC of the lignin, fructan or tannin compositions of a plant; (7) producing a  
 CC plant having one or more of the lignin, fructan or tannin compositions;  
 CC and (8) modifying the activity of (II) involved in a lignin, fructan or  
 CC tannin biosynthetic pathway in a plant. (I) can be used for modulating  
 CC the biosynthesis of lignin, fructan or tannin in a plant. The present  
 CC sequence is used in the exemplification of the present invention.  
 XX  
 XX Sequence 2006 BP; 405 A; 630 C; 615 G; 356 T; 0 U; 0 Other;  
 SQ  
 Query Match 17.0%; Score 85.4; DB 9; Length 2006;  
 Best Local Similarity 51.5%; Pred. No. 4.1e-06;  
 Matches 253; Conservative 0; Mismatches 226; Indels 12; Gaps 2;  
 QY 2 TGTTCTACGTCGGGACGACCGGGGGCCCAAGGGGCTGTCTCCACGACGGCACT 61  
 Db 706 TGGCCTACTCTCCGCGACACCGGGGTCCTCCAAAGGGGCTCATCTACCCACCGCAGCC 765  
 QY 62 GCCTCTGGTGGTGGC-----TTCTCTACGTGCTGCTTCCCGGGTGTGCGACC 112  
 Db 766 TCGTACACAGGCTCGCCAGCTGGTTGACGGGTGGAACCAAGTGTGCTTACACAGG 825  
 QY 113 AGGACCGGGTCTCTGGCGCTCCCGCTGTTCCACAGCTTTTCGCACATCGCTGGCTCC 172  
 Db 826 ACAGCGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885  
 QY 173 TGTCCGACCGTGTGCTGGGCGAGCTGTCGGATCGCGACGGAGTTCGCGCCGACGAGC 232  
 Db 886 TGGCGGGGCTCCGGTCTGGCGCGCGCCATCGTCATCATCGCAAGTTCGACGTGCGCGCGC 945  
 QY 233 TGATCGGCTGATCGAGCGGAGAGCTCGAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 292  
 Db 946 TGGTGGACCTGCTCGCGCGCGCACCGCATCCATCGCGCCATTCGTGCGCGCCATCGTGG 1005  
 QY 293 ACCACCTGGTGGCG 352  
 Db 1006 TGGAGATCGCAAGAGCGACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065  
 QY 353 TGGCGGGGGCGCGCTCTCG 409  
 Db 1066 TCTCCG 1125  
 QY 410 GGGTCCCGCTGATCGACGCTACGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 469  
 Db 1126 ACGCGTGTCTCGGACAGGTTACGGATGACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185  
 QY 470 CGCGCGACGCGC 480  
 Db 1186 TGGCGTTTCCG 1196  
 RESULT 11  
 ID ADC68538  
 XX standard; cDNA; 2014 BP.  
 AC  
 XX ADC68538;  
 XX

DT 18-DEC-2003 (first entry)  
 XX *Lolium perenne* lignin/tannin biosynthesis protein cDNA SEQ ID NO:30.  
 DE  
 XX *Lolium perenne*; *Festuca arundinacea*; lignin; fructan; tannin;  
 KW biosynthetic pathway; plant; gene; ss.  
 KW  
 XX *Lolium perenne*.  
 OS  
 XX WO2003040306-A2.  
 PN  
 XX 15-MAY-2003.  
 PD  
 XX 07-NOV-2002; 2002WO-NZ000239.  
 PF  
 XX 07-NOV-2001; 2001US-0337703P.  
 PR  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (WRIG-) WRIGHTSON SEEDS LTD.  
 XX Denner J, Forster RL, Gibson JB, Shenk MA, Norries MG, Glenn M;  
 PI Saulsbury KW, Hall C;  
 XX WPI; 2003-441544/41.  
 DR P-PSDB; ADC68382.  
 DR  
 XX New polynucleotide encoding polypeptides from *Lolium perenne* or *Festuca*  
 PT arundinacea, useful for modulating the biosynthesis of lignin, fructan or  
 PT tannin in a plant.  
 XX  
 XX Claim 1; SEQ ID NO 30; 240pp; English.  
 PS  
 XX The present invention describes isolated polynucleotides (I) encoding  
 CC proteins (II) from *Lolium perenne* and *Festuca arundinacea* which are  
 CC active in lignin, fructan and tannin biosynthetic pathways. Also  
 CC described: (1) an isolated oligonucleotide probe or primer comprising at  
 CC least 10 contiguous residues complementary to 10 contiguous residues of  
 CC (1); (2) a kit comprising the oligonucleotide probe or primer; (3) a  
 CC genetic construct comprising (1); (4) a transgenic plant cell comprising  
 CC the genetic construct of (3); (5) a plant or its seed, fruit or progeny  
 CC comprising the transgenic plant cell of (4); (6) modulating one or more  
 CC of the lignin, fructan or tannin compositions of a plant; (7) producing a  
 CC plant having one or more of the lignin, fructan or tannin compositions;  
 CC and (8) modifying the activity of (II) involved in a lignin, fructan or  
 CC tannin biosynthetic pathway in a plant. (I) can be used for modulating  
 CC the biosynthesis of lignin, fructan or tannin in a plant. The present  
 CC sequence is used in the exemplification of the present invention.  
 XX  
 XX Sequence 2014 BP; 413 A; 628 C; 618 G; 355 T; 0 U; 0 Other;  
 SQ  
 Query Match 17.0%; Score 85.4; DB 9; Length 2014;  
 Best Local Similarity 51.5%; Pred. No. 4.1e-06;  
 Matches 253; Conservative 0; Mismatches 226; Indels 12; Gaps 2;  
 QY 2 TGTTCTACGTCGGGACGACCGGGGCGGCCCAAGGGGCTGTCTCCACGACGGCAACT 61  
 Db 707 TGGCCTACTCTCTCGGACACCGGGGCTCCCAAGGGGCTCATCTACCCACCGCAGCC 766  
 QY 62 GCCTCTGGTGGTGGC-----TTCTCTACGTGCTGCTTCCCGGGTGTGCGACC 112  
 Db 767 TGGTACACAGCTGCGCCAGCTGGTTGACGGGTGGAACCAAGTGTGCTTCAACAGG 826  
 QY 113 AGGACCGGGTGTCTGGCGCTCCCGCTGTTCCACAGCTTTTCGCACATCGCTGGCTCC 172  
 Db 827 ACAGCGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886  
 QY 173 TGTCCGCGACCGTGTGCGGGGCGAGCTCGGATCGCGACCGGACGCTCCGCGCAGCG 232  
 Db 887 TGGCGGGGCTCGCGTGTGGCGCGCGCATCTGTCATCATGCGCAAGTTCGACGTGCGCGCGC 946  
 QY 233 TGATCGGCTGATCGAGCGGAGAGCTCGAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 292  
 Db 947 TGGTGGACCTGCTCGCGCGCGCGCGCATCCATCGCGCGCGCGCGCGCGCGCGCGCGCG 1006





XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.  
PA (UYAD-) UNIV ADELAIDE.  
PA (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.  
PA (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.  
PA (UYSC-) UNIV SOUTHERN CROSS.  
PA (DAIR-) DAIRY RES & DEV CORP.  
XX Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;  
XX WPI; 2002-097993/13.  
XX P-PSDB; AAU75085.  
XX Novel nucleic acid encoding enzymes involved in lignin biosynthetic  
PT pathway from ryegrass or fescue species useful for modifying lignin  
PT biosynthesis in plants and as a molecular genetic marker.  
XX Claim 3; Fig 3; 148pp; English.  
XX This invention represents purified or isolated nucleic acid and protein  
CC sequences of enzymes involved in lignin biosynthesis. The enzymes of the  
CC invention are 4 coumarate CoA-ligase (4CL), cinnamoyl-CoA reductase (CCR)  
CC and cinnamyl alcohol dehydrogenase (CAD) from a ryegrass (Lolium sp.) or  
CC fescue (Festuca sp.). The invention also comprises an isolated regulatory  
CC element from the nucleic acid sequences and a plant cell or seed  
CC transformed with the nucleic acid. An isolated regulatory element from  
CC these nucleotide molecules is useful for expressing an exogenous gene in  
CC plant cells. The nucleotide sequences of the invention and vectors  
CC containing these sequences are useful for modifying lignin biosynthesis  
CC in a plant and are useful as a molecular genetic marker for qualitative  
CC trait loci (QTL) tagging, QTL mapping, DNA fingerprinting and in marker  
CC assisted selection, in forage and turf grass improvement, e.g. tagging  
CC QTLs for dry matter digestibility, herbage quality, drought tolerance,  
CC regrowth after cutting and grazing, cold tolerance, drought tolerance,  
CC tiller survival and plant persistence. The present sequence represents  
CC the perennial ryegrass 4-coumarate Co-A-ligase 2 (Lp4CL2) cDNA of the  
CC invention  
XX  
XX Sequence 1992 BP; 409 A; 622 C; 576 G; 385 T; 0 U; 0 Other;  
SQ  
Query Match 16.4%; Score 82.2; DB 6; Length 1992;  
Best Local Similarity 51.1%; Pred. No. 1.4e-05;  
Matches 251; Conservative 0; Mismatches 228; Indels 12; Gaps 2;  
QY 2 TGTTCACAGTCGGGACGACCGCGCGCCCAAGGCGTCGTCTCCACGACGCGCACT 61  
Db 636 TGCCGTAATCGTCGGGACGACCGCGCGCTGCCAAGGCGTCATGCTCACCCACGCGAC 695  
QY 62 GCCTCTGGTCCGTGCTTC-----CTGCTACGTGCGGTTCCCGGGTTGCGAAC 112  
Db 696 TCCTACACGCGTCGCGCCACGAGTGACGCGTGAGAACCCGACCTGCACTTACGTCGT 755  
QY 113 AGGACCGGGTGTCTGTGCGCGCTCCCGCTGTTCCACAGCCTTTCCGACATCGCTCGTCC 172  
Db 756 CGACGCTGCTGTGTGGTGTGCTGCGGTGTTCCACATCTACTGCTCACTCGTGTGTCG 815  
QY 173 TGTTCGCGCACCGTGTGCGGCGCACGCTTCGAGATCGCGAGCGGAGCTTCGCGACGACG 232  
Db 816 TCGCGGCTCTCCGCGCGCGGTGCGCATCGTGATCATGCGCAAGTTCGACCCACGCGCGC 875  
QY 233 TGATCGCGCTGATGAGCGGAGAGCTCGACCTTCTCGCGCGGCTGCGACCACTTACC 292  
Db 876 TGTGTGACCTTGGTGGCGACGCGCTCACCTGTGGCGCCATTCTGTGCGCGCCATCGTGG 935  
QY 293 ACCACCTGTTGCGGCGCGCGCGCGCGGCTTTCTCGCGCGCGAGCTTCGCGATCGGCC 352  
Db 936 TGGAGATGCGCAAGAGCGCGGCTGACCGCGCGGACCTGGCTCCATCGCGTGTGCA 995  
QY 353 TGGCGCGGCGCGGCTGCTCGCGCGCGGCTGCGAGCGAGTTC---GAAGACCTTCG 409  
Db 996 TGTTCGCGGCG 1055  
QY 410 GGTTCCTCGATCGACGCTACGCGCAGCGGACCGAGACCTTCGCGGCGCGATCACCATGA 469

Db 1056 ACGCCGTGCTCGGCAGGATATGGATGACCGAGCGCGCCTGTGCTGGCGATGTGCC 1115  
QY 470 CGCCGCGACGGC 480  
Db 1116 TGGCCTCGCC 1126  
RESULT 14  
ABK13573  
ID ABK13573 standard; cDNA; 2038 BP.  
XX  
XX AC ABK13573;  
XX 23-APR-2002 (first entry)  
XX  
XX Ryegrass 4-coumarate Co-A-ligase 3 (Lp4CL3) cDNA.  
XX Perennial ryegrass; ss; lignin; 4 coumarate CoA-ligase; QTL;  
XX lignin biosynthesis; enzyme; cinnamoyl-CoA reductase; CCR; Lp4CL3;  
XX cinnamyl alcohol dehydrogenase; CAD; molecular genetic marker; gene;  
XX qualitative trait loci; tagging; QTL mapping; DNA fingerprinting;  
XX marker assisted selection; forage improvement; turf grass improvement;  
XX dry matter digestibility; herbage quality; palatability; regrowth;  
XX cold tolerance; drought tolerance; tiller survival; plant persistence.  
XX Lolium perenne.  
XX  
XX Key Location/Qualifiers  
FH 5'UTR 1. .112  
FT /tag= a  
FT CDS 113. .1786  
FT /tag= b  
FT /product= "Lp4CL3 protein"  
FT 3'UTR 1787. .2019  
FT /tag= c  
FT polyA\_site 2020. .2038  
FT /tag= d  
XX  
XX WO200195702-A1.  
XX  
XX 20-DEC-2001.  
XX  
XX 14-JUN-2001; 2001WO-AU000699.  
XX  
XX 14-JUN-2000; 2000AU-00008154.  
XX  
XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.  
XX (UYAD-) UNIV ADELAIDE.  
XX (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.  
XX (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.  
XX (UYSC-) UNIV SOUTHERN CROSS.  
XX (DAIR-) DAIRY RES & DEV CORP.  
XX  
XX Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;  
XX WPI; 2002-097993/13.  
XX P-PSDB; AAU75085.  
XX Novel nucleic acid encoding enzymes involved in lignin biosynthetic  
PT pathway from ryegrass or fescue species useful for modifying lignin  
PT biosynthesis in plants and as a molecular genetic marker.  
XX Claim 3; Fig 4; 148pp; English.  
XX This invention represents purified or isolated nucleic acid and protein  
CC sequences of enzymes involved in lignin biosynthesis. The enzymes of the  
CC invention are 4 coumarate CoA-ligase (4CL), cinnamoyl-CoA reductase (CCR)  
CC and cinnamyl alcohol dehydrogenase (CAD) from a ryegrass (Lolium sp.) or  
CC fescue (Festuca sp.). The invention also comprises an isolated regulatory  
CC element from the nucleic acid sequences and a plant cell or seed  
CC transformed with the nucleic acid. An isolated regulatory element from  
CC these nucleotide molecules is useful for expressing an exogenous gene in



Qy 470 CGCCGACGGC 480  
| | | | |  
Db 1197 TGGCGTTGCC 1207

Search completed: June 16, 2004, 23:46:37  
Job time : 290 secs

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	184.2	36.8	77536	4	US-09-410-551B-1	Sequence 1, Appli
C 2	84.2	16.8	1584	4	US-09-252-991A-14523	Sequence 14522, A
C 3	84.2	16.8	1806	4	US-09-252-991A-14654	Sequence 14654, A
C 4	84.2	16.8	1865	4	US-09-252-991A-15096	Sequence 15096, A
C 5	80.8	16.1	1749	4	US-09-489-039A-805	Sequence 805, App
C 6	78.4	15.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 7	78.4	15.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 8	76.6	15.3	30001	1	US-08-125-468-1	Sequence 1, Appli
C 9	76.6	15.3	30061	2	US-08-474-933-1	Sequence 1, Appli
C 10	69.2	13.8	1886	4	US-09-252-991A-3821	Sequence 3821, Ap
C 11	69.2	13.8	2082	4	US-09-252-991A-3777	Sequence 3777, Ap
C 12	69.2	13.8	2181	4	US-09-252-991A-3690	Sequence 3690, Ap
C 13	66	13.2	2124	4	US-09-266-965-44	Sequence 44, Appl
C 14	66	13.2	53500	4	US-09-266-965-76	Sequence 76, Appl
C 15	61.6	12.3	2043	4	US-09-615-192A-266	Sequence 266, App
C 16	61.6	12.3	5103	4	US-09-252-991A-5192	Sequence 5192, Ap
C 17	61.6	12.3	6876	4	US-09-252-991A-5283	Sequence 5283, Ap
C 18	61	12.2	8051	2	US-08-576-628A-2	Sequence 2, Appli
C 19	60.2	12.0	957	4	US-09-252-991A-9229	Sequence 9229, Ap
C 20	60.2	12.0	1182	4	US-09-252-991A-9062	Sequence 9062, Ap
C 21	60.2	12.0	1209	4	US-09-252-991A-8718	Sequence 8718, Ap
C 22	59.4	11.9	549	4	US-09-615-192A-197	Sequence 197, App
C 23	59.2	11.7	357	4	US-09-489-039A-706	Sequence 706, App
C 24	58.8	11.7	1152	4	US-09-252-991A-7729	Sequence 7729, Ap
C 25	58.8	11.7	1215	4	US-09-252-991A-7837	Sequence 7837, Ap
C 26	58.8	11.7	1581	4	US-09-252-991A-7755	Sequence 7755, Ap
C 27	58.2	11.6	1320	4	US-09-252-991A-3999	Sequence 3999, Ap

Db 45901 CTACAGTTCGGCGCTCCGCGCTATCTCTGGGGGCTTCGTCCGGGTGACGCTCTCGAC 45842  
QY 241 CTGATCAGGCGGAGAGCTCGACTTCCTCTGGCGCGGTGCGGACCACTACCAACCTG 300  
Db 45841 GCCCTGCGCGAACTCCAGACCGGTCTGCTGGGGTCCGACGATGTACCGCTACC-- 45784  
QY 301 GTGCGGGCGCGCGGAGCGCGGTTCCTCGCGCGAGCTCGGGATCGGCTTGGCGGG 360  
Db 45783 -TCTCGCGCGCGGTTCCTCGCGCGAGCGCGGCTCGCGCGCTGCGGTGGCTGGTCC 45725  
QY 361 GGGCGGTCTCTCGCGCGCGGCTCGAAGCGAGTTCGAAGAGACCTTCGGGGTCCGCGTG 420  
Db 45724 GGATCCACCTCGCGCGGTCTCTCACAGGATTCGAGCGGAGGTTCGGCGTCCGCTG 45665  
QY 421 ATGACGCCCTACGGCAGCAGGAGCTCGGGGCGGATACCATGAACCGCGCGGAGCG 480  
Db 45664 CTCGACACGTACGGCTGACCGGAGACCGGCTCGCTACCGCCACACACCTGGAGAT 45605  
QY 481 GCCCGGTCTGAGGCTCGTGC 501  
Db 45604 GCGGGGTTCGGGCTCGTGC 45584

## RESULT 2

US-09-252-991A-14522  
; Sequence 14522, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14522  
; LENGTH: 1584  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14522

Query Match 16.8%; Score 84.2; DB 4; Length 1584;  
Best Local Similarity 49.9%; Pred. No. 2.1e-08;  
Matches 240; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 7 TACACGTCGGGACAGCGGCGCGCTCGGTTCCTCGCGGTTCGCGGCGGCTTCGCGGCTGCTC 66  
Db 686 TACACCTCCGGTACACCGGCTTCCCAAGGCGCCACCTCAGCCACTACACATCTC 745  
QY 67 TGGTCCGTTCCTCTGCTAGCTGCGGTTCCTCGCGGTTCGCGGCGGCTGCTC 126  
Db 746 AACACGCTACATGTTGCGGGAAGCTCGGACTCGGCGGGAAGACCGCTGCTGATC 805  
QY 127 TGGCGCTCCGCTGTTCCACAGCCTTCGACATGCGCTGCTGCTGCTGCTGCTGCTGCTG 186  
Db 806 CCGGTCCGCTGTATCACTGCTTCGGCATGTTGATGAGCAACCTCGGCTGCTCACCAC 865  
QY 187 GTCGGGGCAGCGTCGGATCGCGAGCGGAGCTCGCGGCGGAGCTGATCGGCTGATC 246  
Db 866 GGCTCGACCATGATCTACCGCGCGGCTGCTGAGCGGCGGAGCGGAGCGGAGCGGCTG 925  
QY 247 GAGCGGAGAGCTGACCTTCTTGGCGCGGCTGCGGACCACTACCACTGCTGCGG 306  
Db 926 GCCGAGGAGCGCGCACCGCGCTGTACGGGTACCGACCATGTTTCATCGCGGAGCTGGAT 985  
QY 307 GCCGCGGAGCGGCGGTTCCTCGCGCGGCTGCGGCGGAGCTGCGGATCGG--CCTGCGGCGG 363  
Db 986 CATCCGCGCGCGGAAATTCGACCTTTCCAGCTGCGCACCAGGAATCATGGCGCGGCGCC 1045

QY 364 GGGTCTCTCGCGCGGCTCGAAGCGAGTTCGAAGAGACCTTCGGGGTCCGCTGATC 423  
Db 1046 ACCTGCCGATCAGGTGATGCGCGGCTCATTTGGGACATGACATGCGCGAGGTGCG 1105  
QY 424 GAGCGCTACGCGACGACCGGAGCTTCGCGGGGCGATCACCATGAACCGCGCGGCGCC 483  
Db 1106 ATCGCTACGGCATGACCGAGACCACTCCGCTGCTGCTGACAGACCGGTTCGGACGATGGC 1165  
QY 484 C 484  
Db 1166 C 1166

## RESULT 3

US-09-252-991A-14654  
; Sequence 14654, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14654  
; LENGTH: 1806  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14654

Query Match 16.8%; Score 84.2; DB 4; Length 1806;  
Best Local Similarity 49.9%; Pred. No. 2.1e-08;  
Matches 240; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 7 TACACGTCGGGACAGCGGCGCGCTCGGTTCCTCGCGGTTCGCGGCGGCTTCGCGGCTGCTC 66  
Db 733 TACACCTCCGGTACACCGGCTTCCCAAGGCGCGCACCTCAGCCACTACACATCTC 792  
QY 67 TGGTCCGTTCCTCTGCTAGCTGCGGTTCCTCGCGGTTCGCGGCGGCTGCTC 126  
Db 793 AACACGCTACATGTTGCGGGAAGCTTCGACTCGGCGGGAAGACCGCTGCTGATC 852  
QY 127 TGGCGCTCCGCTGTTTCACAGCCTTCGACATGCGCTGCTGCTGCTGCTGCTGCTGCTG 186  
Db 853 CCGGTCCGCTGATCACTGCTTCGGCATGTTGATGCGCAACCTCGGCTGCTCACCAC 912  
QY 187 GTCGGGCGAGCTCGGATCGCGAGCGGAGCTTCGCGGCGGAGCTGATGCGGCTGATC 246  
Db 913 GGCTCGACCATGATCTACCGCGGCGCGGCTTCGACCGCGGAGCGACCTGCTGCGCG 972  
QY 247 GAGCGGAGAGCTGACCTTCCTGGCGCGGCTGCGGACCACTACCACTGCTGCTGCTG 306  
Db 973 GCCGAGGAGCGCGCACCGGCTGTACGGGTACCGACCATGTTTCATCGCGGAGCTGGAT 1032  
QY 307 GCCCGCGGCGGCGGCTTTCCTCGCGCGGAGCTGCGGATCGG--CCTGGCGGCGGCG 363  
Db 1033 CATCCGCGCGCGGAAATTCGACCTTTCCAGCTTGGCGGCTGCGCACCGGAATCATGGCGCGGCGCC 1092  
QY 364 GCGGTCTCTCGCGCGGCTGCGAGCGAGTTCGAAGAGACCTTCGGGGTCCGCTGATC 423  
Db 1093 ACTTCCCGCATCGAGGTGATGCGCGGCTCATTTGGCGACATGACATGCGCGGAGTCCAG 1152  
QY 424 GAGCGCTACGCGAGCACCAGAGCTCGGGGCGGATCACCATGAACCGCGCGGAGCGGCG 483  
Db 1153 ATCGCTTACGCGATGACCGGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1212  
QY 484 C 484

Db 1213 C 1213

RESULT 4  
US-09-252-991A-15096/c  
; Sequence 15096, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: 107196.136  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15096  
; LENGTH: 1965  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15096

Query Match 16.8%; Score 84.2; DB 4; Length 1965;  
Best Local Similarity 49.9%; Pred. No. 2.1e-08;  
Matches 240; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

Qy 7 TACACGTGGGCAAGACCGGGGCGCCCAAGGGGTCGTGTCACGACGCGCAACTGCTC 66  
Db 1305 TACACCTCGGTATACACCGGCTTCCCAAGGGGCGCCACCTCAGCCACTACCAATCCTC 1246

Qy 67 TGGTCCGTGCTCTCTCTAGTCGCGTTCCTCCAGAGCTTTCCCGGGTGTTCGGACCGAGCGGGTCTC 126  
Db 1245 AACAAAGGCTACATGTTGTCGGCGAAAGCCTCGGACTCGGCGCGGAAGACCGCTGTGTGATC 1186

Qy 127 TGGCGGCTCCCGCTGTTCCACAGCCTTTTCGACATCGCTGCTCTCTTCGCGCACCGCTG 186  
Db 1185 CCGTGGCGCTGTATCACTGTTTGGCATGTGTATGGCAACCTCGGTGCTGTATCCAC 1126

Qy 187 GTCGGGGCCAGCTCCGATCCGACCGGACGCTCCGCGACGACGATGATGGGTGATC 246  
Db 1125 GGTTCGACCATGATACCTCCGGGCGCCAGCTTCGACGCGGAGGACCTGCTGGCCGCTG 1066

Qy 247 GAGCGGAGAGCTGACCTTCTGCGCGGCTGCGGACCACTTACCACTGCTGTGGG 306  
Db 1065 GCCGAGGAGCGCCACCGCGCTGTACCGGGGTACCGGGGTACCGACCATGTTTCATCGCGGAGCTGGAT 1006

Qy 307 GCCGCGCGGACGCGGCTTCTCCGCGCGGAGCCTTCGCGATCGG---CTGGCGCGGGGC 363  
Db 1005 CATCGCGCGCGCGGATTCGACCTTTCCAGCTTCGCGACCGGATCATGGCGCGGCC 946

Qy 364 CGGTCTCGGCGCGGCTCGAAGCGAGTTTCGAAGAGACTTCGGGGTCCCGCTGATC 423  
Db 945 ACCTCCCGATCGAGGTGATGCGGGGCTCATTTGGCGACATGCATGCGCGAGGTGCG 886

Qy 424 GACGCTAGGACGACCGGACCTTCGGGGGATCACCATGAACCCGCGGACCGGCC 483  
Db 885 ATCGGCTAGGACGACCGGACCTTCGGGGGATCACCATGAACCCGCGGACCGGCC 826

Qy 484 C 484  
Db 825 C 825

RESULT 5  
US-09-489-039A-805  
; Sequence 805, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 805  
; LENGTH: 1749  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-805

Query Match 16.1%; Score 80.8; DB 4; Length 1749;  
Best Local Similarity 49.0%; Pred. No. 9.5e-08;  
Matches 245; Conservative 0; Mismatches 252; Indels 3; Gaps 1;

Qy 2 TGTTCATACGTGGGCAAGACCGGGGCGCCCAAGGGGTCGTGTCACGACGCGCAACT 61  
Db 665 TCTGTTACCTCGGAAACACACCGCGCGCGGCGAAAGGGGTATTGAACACCATCTTCCG 724

Qy 62 GCCTCTGTCGCTGCTTCTCTAGTTCGCG---TTCCCGGGTTCGACACGAGGACC 118  
Db 725 TGGTCTCTGCAACGCTTGAAGCGCAACAGCGCGGACGCGCGGCATATCGGCGAGGACA 784

Qy 119 GGGTGTCTTGGCGGCTCCCGCTGTTCCACAGCCTTTCGCACATCGCCTGCTGCTGTCG 178  
Db 785 GCCTGTTCCCGGTGTCGCGATGTTCCATGTCACGCGTGGGGGACGCCGTTTATCGCG 844

Qy 179 CCACGCTGTCGGGGCCAGCTTCGCGATTCGCGGACGCGGACGCTCCGCGCACGACGTGATGC 238  
Db 845 CCATGCTCGGCGCGGCTGCTGTCGCGGGCGCGCACCTCGATGCGCGACAGCCTGCTGC 904

Qy 239 GGCTGATCGAGCGGACGCTGACCTTCTTCGCGCGGCTCCGCGGACGCGCACCTACCAACC 298  
Db 905 AGCTGCTGCGCGCGGAGAGGTGACGCTGGGGTTCGCGGCTGCTGCGGCGGCT 964

Qy 299 TGGTGGCGGCGCGCGCGGCGGCTTCTTCGCGCGGAGCTGCGGATCGCCTGCGCG 358  
Db 965 TGCTGGCGGCGATGCGCGGACAGAGTTCGCGTTCCTGATTTAAAGTGGCTGGTTCG 1024

Qy 359 GGGGCGCGGCTCTCGCGCGCGGCTGCGAAGCGAGTTTCGAAGAGACCTTCGGGGTCCCGC 418  
Db 1025 GCGGCTCGCGCTGCGCGCGTGCATGGCGGAGGCTTTCAGCGTGATTACGCGATTGCAC 1084

Qy 419 TGATCGAGCCTACGCGACGACCGGACCTTCGCGGCGGATCACCATGAACCGCGGAGC 478  
Db 1085 TGACCCAGCCTGGGGAATGACCGAAACATCGCCCATCGGCACCATCACACCGTTGA 1144

Qy 479 GCGCGCGGCTCGAGGGCTCG 498  
Db 1145 GCBAACATGACGCGCTCCG 1164

RESULT 6  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765





```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match          15.3%; Score 76.6; DB 1; Length 30001;
Best Local Similarity 52.7%; Pred. No. 6.4e-07;
Matches 166; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 176 CCGCCACCGTGTGCGGGCCAGCGTCGGAGTCGCGGAGCGGCGAGCTCCGCCGACGACGTGA 235
DB 9748 CCGCGCTCGCGGTGCGGCGGAGCGACGCTGTCTGCCGAGAGAGACCGCGCGGCGTGG 9807

QY 236 TGGGCTGATCGAGCGGAGAGCTCGACCTTCTCTGGCGCGGCTGCGGACCTTACACACC 295
DB 9808 TGGGACTGCGCGGTACCTTGGCGGCGCTCGACCTGTACAGCTGCGCGGTGCTGTCCC 9867

QY 296 ACCTGGTGGCGGCGCGCGCGGCGGCTTCTCGCGCGGAGCTGCGGAGCTGCGGCTGG 355
DB 9868 GGCTGCGCGGAGAGCGGCTGCGCGGCTTCCGATCCCTCGCTACGCGGCGGTGCTGT 9927

QY 356 CCGGGGCGCGGCTCTCGCGCGCGGCTGCGAGCGAGTTCGAAGAGACCTTCGGGGTCC 415
DB 9928 CCGGTGGATCGCGCTGCGCGCGCGGCTGCGAGCGAGCTGCGGCGGCTTCCGCGGTGC 9987

QY 416 CGCTGATCGAGCTTACGCGAGCACCAGAGACCTGCGGGGCGATCACCATGACCGCGCG 475
DB 9988 CGGTGCTCCAGGCTTACGCGCTGCGCGGAGAGCTGCCCGTCCGAGACCTTCGACCTGCCCCG 10047

RESULT 9
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match          15.3%; Score 76.6; DB 2; Length 30001;
Best Local Similarity 52.7%; Pred. No. 6.4e-07;
Matches 166; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 176 CCGCCACCGTGTGCGGGCCAGCGTCGGAGTCGCGGAGCGGCGAGCTCCGCCGACGACGTGA 235
DB 9748 CCGCGCTCGCGGTGCGGCGGAGCGACGCTGTCTGCCGAGAGAGACCGCGCGGCGTGG 9807

QY 236 TGGGCTGATCGAGCGGAGAGCTCGACCTTCTCTGGCGCGGCTGCGGACCTTACACACC 295
DB 9808 TGGGACTGCGCGGTACCTTGGCGGCGCTCGACCTGTACAGCTGCGCGGTGCTGTCCC 9867

QY 296 ACCTGGTGGCGGCGCGCGCGGCGGCTTCTCGCGCGGAGCTGCGGAGCTGCGGCTGG 355
DB 9868 GGCTGCGCGGAGAGCGGCTGCGCGGCTTCCGATCCCTCGCTACGCGGCGGTGCTGT 9927

QY 356 CCGGGGCGCGGCTCTCGCGCGCGGCTGCGAGCGAGTTCGAAGAGACCTTCGGGGTCC 415
DB 9928 CCGGTGGATCGCGCTGCGCGCGCGGCTGCGAGCGAGCTGCGGCGGCTTCCGCGGTGC 9987

QY 416 CGCTGATCGAGCTTACGCGAGCACCAGAGACCTGCGGGGCGATCACCATGACCGCGCG 475
DB 9988 CGGTGCTCCAGGCTTACGCGCTGCGCGGAGAGCTGCCCGTCCGAGACCTTCGACCTGCCCCG 10047

RESULT 10
US-09-252-991A-3821
; Sequence 3821, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3821
; LENGTH: 1986
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3821

Query Match          13.8%; Score 69.2; DB 4; Length 1986;
Best Local Similarity 55.8%; Pred. No. 1.7e-05;
Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;

QY 225 CGACGACGTGATGCGGCTGATCGAGCGGAGAGCTTCGACCTTCCTGGCGCGGCGTCCGAC 284
```

Db 987 CCAGGACTTCTGGAAGCTGTTGGAACGCTACCGGGTCAACAGCTTTCAGTGGCGTCCGAC 1046  
 QY 285 CACCTACACACCTGCTGCTGGCGGCGCGCGGAGCGGGTTCCTCGCGCGGAGCTGCG 344  
 Db 1047 CATCTACGCGGGCTGCTGCTGAGTCCCGAGCGATGCGCG---CGACCTGTCCAGCCTGG 1103  
 QY 345 GATCGGCTGCGCGGGCGCGGCTGCTGCGGCGCGGCTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 1104 TTTGCGCCCTGTGGCGCGCGCGCGGATGCGGTGAGCTGATCCGCGAGTTCAGGCGCG 1163  
 QY 405 CTTTGGGGTCCCGCTGATCGACGCTTACGCGACACCGAGACCTGCGGGCGGATCACCAT 464  
 Db 1164 TACCGGGCTCAAGTGTATCGAGGCTTACGCGCTCACCGAAGGTACCTGCGGCACCAATTG 1223  
 QY 465 GAACCGCGCGAGCGCGCGCGCTGCGAGGCTCG 498  
 Db 1224 CAACCGCGTGGCGCGAGCGCGCGGCTTTCG 1257

RESULT 11  
 US-09-252-991A-3777  
 ; Sequence 3777, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 3777  
 ; LENGTH: 2082  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-3777

Query Match 13.8%; Score 69.2; DB 4; Length 2082;  
 Best Local Similarity 55.8%; Pred. No. 1.7e-05;  
 Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;  
 QY 225 CGAGGAGCTGATGCGGCTGATCGAGGCGGAGAGCTCGACCTTCTCGCGCGGCTGCGGAC 284  
 Db 1156 CCAGGAGCTTCTGGAAGCTGTTGGAACGCTACCGGGTCAACAGCTTTCAGTGGCGTCCGAC 1215  
 QY 285 CACCTACACACCTGCTGCTGGCGGCGCGCGGAGCTGCTTCTCGCGCGGAGCTGCG 344  
 Db 1216 CATCTACGCGGCTGCTGCTGAGTCCCGAGTGCGCG---CGACCTGTCCAGCCTGG 1272  
 QY 345 GATCGGCTGCGCGGGCGCGGCTGCTGCGGCGCGGCTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 1273 TTTGCGCCCTGTGGCGCGCGCGGATGCGGTGAGCTATCGCGCAGCTTTCAGGCGCG 1332  
 QY 405 CTTTGGGGTCCCGCTGATCGACGCTTACGCGACACCGAGACCTGCGGGCGGATCACCAT 464  
 Db 1333 TACCGGGCTCAAGTGTATCGAGGCTTACGCGCTCACCGAAGGTACCTGCGGCACCAATTG 1392  
 QY 465 GAACCGCGGAGCGCGCGCGCTGCGAGGCTCG 498  
 Db 1393 CAACCGCGTGGCGCGAGCGCGCGGCTTTCG 1426

RESULT 12  
 US-09-252-991A-3690/c  
 ; Sequence 3690, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 3690  
 ; LENGTH: 2181  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-3690  
 Query Match 13.8%; Score 69.2; DB 4; Length 2181;  
 Best Local Similarity 55.8%; Pred. No. 1.7e-05;  
 Matches 153; Conservative 0; Mismatches 118; Indels 3; Gaps 1;  
 QY 225 CGAGGAGCTGATGCGGCTGATCGAGGCGGAGAGCTCGACCTTCTCGCGCGGCTGCGGAC 284  
 Db 1015 CCAGGAGCTTCTGGAAGCTGTTGGAACGCTACCGGGTCAACAGCTTTCAGTGGCGTCCGAC 956  
 QY 285 CACCTACACACCTGCTGCTGGCGGCGCGCGGAGCTGCTTCTCGCGCGGAGCTGCG 344  
 Db 955 CATCTACGCGGCTGCTGCTGAGTCCCGAGTGCGCG---CGACCTGTCCAGCCTGG 899  
 QY 345 GATCGGCTGCGCGGGCGCGGCTGCTGCGGCGCGGCTGCGAAGCGAGTTCGAAGAGAC 404  
 Db 898 TTTGCGCCCTGTGCGCGCGCGCGGATGCGGTGAGCTATCGCGCAGTTCAGGCGCG 839  
 QY 405 CTTTGGGGTCCCGCTGATCGACGCTTACGCGACACCGAGACCTGCGGGCGGATCACCAT 464  
 Db 838 TACCGGGCTCAAGTGTATCGAGGCTTACGCGCTCACCGAAGGTACCTGCGGCACCAATTG 779  
 QY 465 GAACCGCGGAGCGCGCGCGCTGCGAGGCTCG 498  
 Db 778 CAACCGCGTGGCGCGAGCGCGCGGCTTTCG 745

RESULT 13  
 US-09-266-965-44  
 ; Sequence 44, Application US/09266965  
 ; Patent No. 6495348  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/09/266,965  
 ; CURRENT FILING DATE: 1999-03-12  
 ; EARLIER APPLICATION NUMBER: US 08/624,447  
 ; EARLIER FILING DATE: 1996-08-19  
 ; EARLIER APPLICATION NUMBER: PCT/US94/11279  
 ; EARLIER FILING DATE: 1994-10-06  
 ; EARLIER APPLICATION NUMBER: US 08/133,963  
 ; EARLIER FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 44  
 ; LENGTH: 2124  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces lavendulae  
 US-09-266-965-44

Query Match 13.2%; Score 66; DB 4; Length 2124;  
 Best Local Similarity 49.6%; Pred. No. 7.2e-05;  
 Matches 233; Conservative 0; Mismatches 225; Indels 12; Gaps 2;  
 QY 2 TGTCTACAGTCGCGGACGACCGCGGCGCCCAAGGCGTGTGTCCACGAGCGCAACT 61



QY	297	CCTGGTGGGGCCCGCCGCGGCGGTTCTCCGCGCGAGCCTCGGATCGGCCTGGC	356
Db	949	GATCGCCAGAGCGCCGAGGTGGACCGGTACGACCTGTCTCGATCCGACCATCATGTC	1008
QY	357	CGGGGGCGGGTCTCTCGGCGCCGGGT---GCCAAGCGAGTTGGAAGAGACCTTCGGGGT	413
Db	1009	GGGTGCGGCCCGGATGGGGAAGGAGCTCGAGGACACCGTCCGAGCCAAGCTGCCCAATGC	1068
QY	414	CCGCTGATCGACGCTACGGGAGCACCAGACCTCGGGGGCGATCACCATGAACCGGC	473
Db	1069	CAAGCTCGGACAGGGGTATGGGATGACGGAGCGGGCCCGGTGCTGCAATGTGCCCGGC	1128

Search completed: June 17, 2004, 03:21:44  
 Job time : 68.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds

(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 agttctacacgtcgggcac.....ccgcgtcgaggctcgatgc 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications NA:\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
  - 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
  - 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
  - 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
  - 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
  - 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	285	56.9	692	17	US-10-461-194-11 Sequence 11, Appl
2	285	56.9	85692	17	US-10-461-194-1 Sequence 1, Appl
3	283.4	56.6	86941	17	US-10-461-194-2 Sequence 2, Appl
4	260.2	51.9	644	17	US-10-461-194-23 Sequence 22, Appl
5	258.6	51.6	644	17	US-10-461-194-22 Sequence 23, Appl
6	221.8	44.3	405	17	US-10-461-194-32 Sequence 32, Appl
7	184.2	36.8	77536	10	US-09-940-316B-1 Sequence 1, Appl
8	114	22.8	1497	16	US-10-369-493-32037 Sequence 32037, A
9	109.6	21.9	1500	16	US-10-369-493-43198 Sequence 43198, A
10	107.8	21.5	1566	16	US-10-369-493-42307 Sequence 42307, A
11	101.2	20.2	1485	15	US-10-156-761-373 Sequence 373, Appl
12	101.2	20.2	9025608	15	US-10-156-761-1 Sequence 1, Appl
13	98.8	19.7	2520	15	US-10-149-485-1 Sequence 1, Appl
14	96.2	19.2	1455	15	US-10-156-761-1335 Sequence 1335, Ap

c	15	96.2	19.2	9025608	15	US-10-156-761-1	Sequence 1, Appl
	16	92.8	18.5	1509	15	US-10-156-761-1592	Sequence 1592, Ap
	17	90.8	18.1	1542	16	US-10-369-493-44547	Sequence 44547, A
	18	90.2	18.0	1478	16	US-10-260-238-3996	Sequence 3996, Ap
	19	87.8	17.5	1584	13	US-10-282-122A-11150	Sequence 11150, A
	20	87.8	17.5	18876	15	US-10-329-079-42	Sequence 42, Appl
	21	87.8	17.5	61944	15	US-10-329-079-34	Sequence 34, Appl
	22	86.8	17.3	1572	16	US-10-369-493-28305	Sequence 28305, A
	23	86.8	17.3	1572	16	US-10-369-493-31062	Sequence 31062, A
	24	86.6	17.3	1527	16	US-10-369-493-31819	Sequence 31819, A
	25	85.4	17.0	2006	15	US-10-289-757-143	Sequence 143, App
	26	85.4	17.0	2014	15	US-10-289-757-30	Sequence 30, Appl
	27	84.8	16.9	1467	16	US-10-369-493-28196	Sequence 28196, A
	28	84.8	16.9	1467	16	US-10-369-493-30954	Sequence 30954, A
	29	84.2	16.8	1482	16	US-10-369-493-41560	Sequence 41560, A
	30	82.8	16.5	1473	16	US-10-369-493-31547	Sequence 31547, A
	31	82.4	16.4	1366	13	US-10-425-114-25299	Sequence 25299, A
	32	82.2	16.4	2073	15	US-10-289-757-31	Sequence 31, Appl
	33	81.4	16.2	7788	15	US-10-329-079-8	Sequence 8, Appl
	34	81.4	16.2	37360	15	US-10-329-079-6	Sequence 6, Appl
	35	81.2	16.2	1548	16	US-10-369-493-28210	Sequence 28210, A
	36	81.2	16.2	1548	16	US-10-369-493-30969	Sequence 30969, A
	37	80.6	16.1	2006	15	US-10-182-113A-2	Sequence 2, Appl
	38	80.6	16.1	2075	13	US-10-425-114-889	Sequence 889, App
	39	80.6	16.1	2122	15	US-10-361-460-20	Sequence 20, Appl
	40	79.4	15.8	1527	16	US-10-369-493-35934	Sequence 35934, A
	41	78.6	15.7	1855	15	US-10-289-757-27	Sequence 27, Appl
	42	78.6	15.7	1470	16	US-10-369-493-32475	Sequence 32475, A
	43	77.4	15.4	2039	15	US-10-289-757-142	Sequence 142, App
	44	77.4	15.4	2041	15	US-10-289-757-28	Sequence 28, Appl
	45	77	15.4	1569	16	US-10-369-493-43103	Sequence 43103, A

## ALIGNMENTS

### RESULT 1

US-10-461-194-11  
; Sequence 11, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Raucher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-GLUTAMINYLIN PRODUCING POLYPEPTIDE SYNTHASES AND  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus





```
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 86941
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-2

Query Match      56.6%; Score 283.4; DB 17; Length 86941;
Best Local Similarity 72.9%; Pred. No. 3.2e-59;
Matches 365; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 1 ATGTTCTACAGTCGGGACGACGCGGGCGCCCAAGGGCGTCTGTCACGACGCGCAAC 60
Db 28184 ATGCTCTACACCTCGGGACACACGGGCGCTCGCGAAAGCGGTGTCTACCCACGCGAAC 28243

Qy 61 TGCCTCTGGTCCGTCGCTTCTGCTAGCTGCGCTTCCCGGGTTTGTCCGACCGAGACCGG 120
Db 28244 TGCCTGTGGTCCCTGGCGGCTGTCTAGTACCGGTGACGGGCTGACCGCGAGGACGGT 28303

Qy 121 GTGCTCTGGCGGCTCCCGCTGTTCACAGCCTTTCCGACATCGCTGCTGCTGCTGCGCC 180
Db 28304 GTGCTGTGGCGGCTGCGGCTGTTCACAGCCTTCCGACATCGTGTGCTGCTGCGCGCC 28363

Qy 181 ACCGTGTGCGGGCCAGCTCCGATCGCGACGCGAGCTCCGCGAGCGTGTGATGCGG 240
Db 28364 ACCGCGGTGCGGGCCGACCGCGATCGTACGCGGGGTGTGACCTCCGATGTCTGAC 28423

Qy 241 CTGATCGAGCGGAGAGCTCGACCTTCTGCGCGGGCGTGGCGACCACTACCAACCTG 300
Db 28424 GCGCTGCGCGAGGAGCGTCTGACCTTCTATCGCGAGTGGCGACGCTCTACCAACCTG 28483

Qy 301 GTGCGGCGCGCGCGCGCGGCTTCTCGCGCGCGAGCTGCGGATCGCGCTGCGCGGG 360
Db 28484 ATCGAGGCTGCGCGAGCGCGAGCTTCCGACGCGCGAGCTGCGGATCGCGCTGCGGGC 28543

Qy 361 GGCGCGCTCTCGCGCGCGGCTCGAGCGAGTTCGAAGAGACTTCGGGGTCCCGCTG 420
Db 28544 GGGCGGTGGCCACGCGCGACCTGGTCAAGTCTGTTGAGGGCCACTTCGAGTGCCTC 28603

Qy 421 ATCGACGCTACGGCAGCACCGAGACCTGCGGGGCGATCACCATGAACCCCGCGAGCGG 480
Db 28604 GTCGACGCTACGGATCCACCGAGACCTGTGCGCGATCGCGGTGAACCTGGGCCACCGG 28663

Qy 481 GCGCGGTGAGGGCTCTGTC 501
Db 28664 CCACGGGTGAGGGTCTGTC 28684

RESULT 4
US-10-461-194-23
; Sequence 23, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GLUTAMINYL PRODUCING POLYKETIDE SYNTHASES AND
; FILE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 3062200700
; CURRENT APPLICATION NUMBER: US/10/461,194
```

```
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-23

Query Match      51.9%; Score 260.2; DB 17; Length 644;
Best Local Similarity 71.8%; Pred. No. 2.5e-53;
Matches 354; Conservative 0; Mismatches 138; Indels 1; Gaps 1;

Qy 9 CAGTCGGGACGACGACGCGGGCGCCCAAGGGCGTCTGTCACGACGCGCAACTGCTCTG 68
Db 1 CGCTCGGGACACACGCGGCGCTGCGAAAGCGTGTCTCCACCCAGCGGAATGCTCTGTG 60

Qy 69 GTCCGCTCGCTTCTGCTACGTTGCGGTTTCCCGGGTTTGTGGACCGAGGACCGGGTCTG 128
Db 61 GTCTCTGCGGCTGTCTAGTACCGGTGACGGGCTGACCGCGAGGACCGTGTGCTGTG 120

Qy 129 GCGCTCGCGGTGTCACAGCCTTTTCGACATCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 188
Db 121 GCGCTCGCGGTGTCGCGGCTTCTGCGACATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 180

Qy 189 CGGGGCGAGGTCGCGGATCGCGACGCGAGCTCCGCGAGCGAGCTGATGCGGCTGATCGA 248
Db 181 CGGGGCGCGACCGCGATCG--TGACGGGCTGCGACCTCCGATGTGCTGAGCGGCTGCG 239

Qy 249 GCGCGAGAGTTCGACCTTCTGCGCGCGGTTCGCGACCACTTACCACTTGTGCTGAGCGG 308
Db 240 CGAGGAGCGGTGCGACCTTCTATCGCGGAGTGCACGCTTACCACTTGTGCTGAGGCG 299

Qy 309 GCGCGGAGCGCGGTTTCTCGCGCGAGCTTCGGGATCGGCTTGGCGGGGGCGGCT 368
Db 300 TGCCTGCGAGCGGACTTTCGCGCACACCGAGCTCGGATCGCTCGTGGGCGGGGCGGT 359

Qy 369 CCTCGGCGCGGGCTCGAAGCGAGTTTGAAGAGACCTTTCGGGTTCCCGCTGATCGAGCG 428
Db 360 GCGCAGCGCGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 419

Qy 429 CTACGCGAGCACCGAGACCTCGCGGGCGATCACCATGAACCCCGCGGACCGCGCGCGGT 488
Db 420 CTACGAGTCCACCGAGACCTTGGCGCGATCGCGTGGACTGCGCGCGCGCGCGCGGCT 479

Qy 489 CGAGGGCTCGTGC 501
Db 480 CGAGGGGTCTATGC 492

RESULT 5
US-10-461-194-22
; Sequence 22, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
```

APPLICANT: Schirmer, Andreas  
APPLICANT: Mcdaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
TITLE OF INVENTION: PRO-GLDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
FILE REFERENCE: 300622009700  
CURRENT APPLICATION NUMBER: US/10/461,194  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130  
PRIOR FILING DATE: 2002-12-13  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 644  
TYPE: DNA  
ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-22

Query Match 51.6%; Score 258.6; DB 17; Length 644;  
Best Local Similarity 71.8%; Pred. No. 6e-53;  
Matches 353; Conservative 0; Mismatches 139; Indels 1; Gaps 1;  
QY 9 CACCTCGGGCAGACCGGGCGGCCCGGCTGCTCCACGCGAGCGCAACTGCTCTG 68  
Db 1 CACCTCGGGCAGACCGGGCGGCCCGGCTGCTCCACGCGAGCGCAACTGCTCTG 59  
QY 69 GTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 128  
Db 60 GTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 119  
QY 129 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 188  
Db 120 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179  
QY 189 CGGGGCGAGGCTCGGATCGCGAGCGGAGCTCGCGCGAGCGATGATGCGGCTGCA 248  
Db 180 CGGGGCGAGGCTCGGATCGCGAGCGGAGCTCGCGCGAGCGATGATGCGGCTGCA 239  
QY 249 GCGGAGAGGCTGACCTTCTGCGCGCGGCTGCGGAGCGGAGCGGAGCGGAGCGG 308  
Db 240 CGAGGAGGCTGACCTTCTGCGCGCGGCTGCGGAGCGGAGCGGAGCGGAGCGG 299  
QY 309 GCGCGGCGAGCGGCTGCTGCGCGCGGCTGCGGAGCGGAGCGGAGCGGAGCGG 368  
Db 300 TGCCCGCGAGCGGAGCTGCGCGCGGCTGCGGAGCGGAGCGGAGCGGAGCGG 359  
QY 369 CCTCGGCGCGGCTGCGGAGCGGAGCTGCGGAGCGGAGCGGAGCGGAGCGGAG 428  
Db 360 GCGCAGCGGCGGCTGCGGAGCGGAGCTGCGGAGCGGAGCGGAGCGGAGCGGAG 419  
QY 429 CTACGCGAGCAGCGGAGCTGCGGCGGAGCTGCGGAGCGGAGCGGAGCGGAGCGG 488  
Db 420 CTACGCGATCCACCGAGAGCTGCGGCGGAGCTGCGGAGCGGAGCGGAGCGGAG 479  
QY 489 CGAGGCTGCTGCG 501  
Db 480 CGAGGCTGCTGCG 492

RESULT 6  
US-10-461-194-32/c  
; Sequence 1, Application US/09940316B

Sequence 12, Application US/10461194  
Publication No. US20040077058A1  
GENERAL INFORMATION:  
APPLICANT: Hutchinson, Richard C.  
APPLICANT: Reid, Ralph C.  
APPLICANT: Hu, Zhihao  
APPLICANT: Rascher, Andreas  
APPLICANT: Schirmer, Andreas  
APPLICANT: Mcdaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
TITLE OF INVENTION: PRO-GLDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
FILE REFERENCE: 300622009700  
CURRENT APPLICATION NUMBER: US/10/461,194  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130  
PRIOR FILING DATE: 2002-12-13  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 32  
LENGTH: 405  
TYPE: DNA  
ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-32

Query Match 44.3%; Score 221.8; DB 17; Length 405;  
Best Local Similarity 72.1%; Pred. No. 4.1e-44;  
Matches 289; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
QY 34 AAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 93  
Db 405 AAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 346  
QY 94 TTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 153  
Db 345 GTGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 286  
QY 154 TGCCACATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 213  
Db 285 TGCCACATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 226  
QY 214 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 273  
Db 225 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 166  
QY 274 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 333  
Db 165 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 106  
QY 334 CCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393  
Db 105 CCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46  
QY 394 TTCCGAGAGAGCTTCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 434  
Db 45 TTCCGAGAGAGCTTCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5

RESULT 7  
US-09-940-316B-1/c  
; Sequence 1, Application US/09940316B



NAME/KEY: gene  
LOCATION: (124)..(987)  
OTHER INFORMATION: ech  
FEATURE:  
NAME/KEY: RBS  
LOCATION: (978)..(981)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (987)..(2459)  
OTHER INFORMATION: fcs, gene of feruloyl-CoA synthetase  
FEATURE:  
NAME/KEY: gene  
LOCATION: (987)..(2462)  
OTHER INFORMATION: fcs  
US-10-149-485-1

Query Match 19.7%; Score 98.8; DB 15; Length 2520;  
Best Local Similarity 54.5%; Pred. No. 8.7e-15;  
Matches 243; Conservative 0; Mismatches 197; Indels 6; Gaps 2;  
QY 2 TGTCTTACACGTCGGGACGACCGGCGCGCCCAAGGGCGTGTCTCCAGCGACGCGCAACT 61  
Db 1438 TGTATGTACACTCGGCGAGCAGCGGACGCGCCCAAGGGCGGATGCTCACCCAGGCAAC 1497  
QY 62 GCTCTGTGTCGGTTCCTGTCTACGTGCGGTTCCTCCGGGTGTTCGACCGAGACCGGG 121  
Db 1498 TCACCTGGAACTCGCTCAACGTCCTGTGTGAGA---CCGACCTGGCGAGCGACGAGCGGG 1554  
QY 122 TGTCTGTGGCGCTCCCGCTGTTCACAGGCTTTTCGACATCGCTCGCTCTCTCCGCCA 181  
Db 1555 CATGTGTGCGCGCGCTGTTCACGCGCGCGCTTCGCGATGTGTGTGCTGTGCTGCTGCGG 1614  
QY 182 CGTGTGTGCGGGCGACGTCGCGATCGCGACGCGGAGCTTCGCGACGAGCTGATGCGGC 241  
Db 1615 TGTCTAAGGGCGGACGCTGATCTTGCACTTCGCGCTTCGACCCCGCGCGTCTGTCCG 1674  
QY 242 TGTATGAGGCGGAGAGCTGCACTTCTGTGCGGGGTTCGCGACCTACCAACCACTGG 301  
Db 1675 CGGTGGAAACGAGAGCGGGTCACGCTGCTGTTCGCGCTGCGCCACGATGTACAGCGCATG 1734  
QY 302 TGGCGGCGCGCGGACGCGGTTCCTCGCGCGGAGCTTCGCGAGCTTCGCGGCTGCGCGGG 361  
Db 1735 CGCGCAACCGCGGTGGCGGAGCGGACCTGTTCAGCTTCGCGACCTCTGTGCGGCG 1794  
QY 362 GCGCGGTTCCTCGCGCGCGGCTGCGAAGCGAGTTTCAAGAGACCTTCGCGGGTTCGCGTGA 421  
Db 1795 GCGCGCGGTTCGCGCGGACCTTCGCGACGCGCTACCTCGA---CCGCGGCTCGCGTTCG 1851  
QY 422 TCGACGCTACGCGAGCACCGGAC 447  
Db 1852 TGCAGGCGCTACGCGCATGACCGAGGCC 1877

RESULT 14  
US-10-156-761-1335  
Sequence 1335, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-05-30  
NUMBER OF SEQ ID NOS: 15109

PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1335  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1455)  
US-10-156-761-1335

Query Match 19.2%; Score 96.2; DB 15; Length 1455;  
Best Local Similarity 51.3%; Pred. No. 4e-14;  
Matches 249; Conservative 0; Mismatches 233; Indels 3; Gaps 1;  
QY 2 TGTCTTACACGTCGGGCGGACCGGCGCGCCCAAGGGCGTGTCTCCAGCGACGCGCAACT 61  
Db 416 TGTATGTACACGAGCGGACGCGGCGCGCCCAAGGGGTGTGTCTCGACCAACGCGCA-- 473  
QY 62 GCTCTGTGTCGGTTCCTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 121  
Db 474 -CTCGCGCGCATGGCGGAGATGATGACCGGCAACCGCGCGGCTGACCGAGACCGACCA 532  
QY 122 TGTCTGTGGCGCTCCCGCTGTTCACAGGCTTTTCGACATCGCTCGCTCTGTCTGTCTGTCT 181  
Db 533 GCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 592  
QY 182 CGTGTGTGCGGGCGACGTCGCGATTCGCGAGCGGAGCTTCGCGCGGAGCTGTGTGTGTGTGT 241  
Db 593 TGTGTGCGCGCGCGCGGTCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 652  
QY 242 TGTATGAGGCGGAGAGCTGCACTTCTGTGCGCGGCTTCGCGACCTACCAACCACTGG 301  
Db 653 TGTGTGCGCGGTTCGCGCGGACCTGTCTTCTGTGCGGTTCGCGGAGTCTACTGTGTGTGT 712  
QY 302 TGGCGGCGCGCGGACGCGGTTCCTCGCGCGGAGCTTCGCGGATCGGCTGTGCGCGGG 361  
Db 713 CGCACTGTGCGGACGAGCTACGCGCGGACCACTCTGTGCGGTACGCTTTCGCGCGGG 772  
QY 362 GCGCGGTTCCTGCGCGCGGCTGCGAAGCGAGTTTCAAGAGACCTTCGCGGGTTCGCGTGA 421  
Db 773 GCGCGCGGATGCGCGCGCGGTCTCATGAGCGGTTCGAGCGCGCTACGACATCCCGGTTC 832  
QY 422 TCGACGCTACGCGACGCGGACCTGTGCGGCGGATCACCATGACCGCGCGGAGCGG 481  
Db 833 TGGAGGCGGTACGCGCTTCCTCGAGGGGACCTTCGCGGCTCCACCAACCGCGCTGTACGG 892  
QY 482 CCGCG 486  
Db 893 GCGCG 897

RESULT 15  
US-10-156-761-1/c  
Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109

```
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      19.2%; Score 96.2; DB 15; Length 9025608;
Best Local Similarity 51.3%; Pred. No. 1e-14;
Matches 249; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

Qy 2 TGTTCACGTCGGGACGACCGGGCGGCCCAAGGGCGTGTCTCACGCGAGCGCAACT 61
Db 1660786 TGATCTACACGAGCGGAGCAACGGGGCGGCCCAAGGGTGTGATCTCGACCAACCCAA-- 1660729

Qy 62 GCCTCTGGTCGGTCGCTTCCCTGCTACGTGCGGTTCCCGGGTGTGCGACCCAGGACCGGG 121
Db 1660728 -CCTCGCGCCCATGGCGGAGATGATGACCGGCAACGCCCGGCTACCGAGACCGACCA 1660670

Qy 122 TGCTCTGCGCCGCTCCCGGTGTTCACAGACCTTTTCGCACATCGCCTGGCTCCTGTCCGCCA 181
Db 1660669 GCCTGCTGATCTCGCGCTGTTCACAGTCAACGGGATCGTGTGCGGGTGTGTCCCGCG 1660610

Qy 182 CCGTGGTCGGGGCCAGCGTCGCGATCGCGACGCGGAGCTCCGCGGACGACGATGATGCGGC 241
Db 1660609 TGCTGGCGGGCGCGCGGTCACGGTGGCGGGCGGTTCCGGGGCGGAGACCTTCTTCGACC 1660550

Qy 242 TGATCGAGGGCGGAGAGCTCGACCTTCTTGGCGGGCTGCGGACCACTTACCACCACTGG 301
Db 1660549 TGGTGGCCACCGTCCGCCCGGACCTGCTTCTCCGGGTCCCGGGGATCTACTCGATGCTCG 1660490

Qy 302 TGGCGGGCGCGCGGAGCGCGGTTTCTCCGGCGCGGAGCTTCGGATCGGCTGGCCCGGGG 361
Db 1660489 CCGAATGCGCGGACCACTGACGCGCGGACACTCTGTCGTTACGGTTGCGCGCTGCGGGG 1660430

Qy 362 GCGCGGTCTCGGGCGCGGGCTGCGAGCGAGTTGGAAGAGACCTTCGGGGTCCCGCTGA 421
Db 1660429 CGGCGCCGATGCCGCGCGCGCTCATCGAGCGGTTTCGAGCGCGGCTACGACATCCCGGTCC 1660370

Qy 422 TCGAGCGCTACGCGAGCACCGGACCTCGGGGGGATCACCATGAACCCGCGGACGCGG 481
Db 1660369 TGGAGGGCTACGGCTCTCCGAGGGGACCTGCGGCTCCACCAACCCGCTGTACGGCC 1660310

Qy 482 CCCGC 486
Db 1660309 GGCGC 1660305
```

Search completed: June 17, 2004, 03:36:44  
Job time : 318.667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:18:23 ; Search time 1960 Seconds

(without alignments)  
7633.139 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 atgttctacgtcgagggcac.....ccgcgtcgagggctcgtgc 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estinu:\*

4: em\_estinu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	18.0	684	13	CA012845 HT06J23r
2	84.8	16.9	564	14	CD865561 AZ02.101D
3	83.6	16.7	666	12	B1958356 HVSME001
4	80.8	16.1	588	13	CA092745 SCVPAM206

5	80.6	16.1	2090	11	AY105108
6	80.4	16.0	543	9	AI665488
7	75.4	15.0	536	14	CD056352
8	74.4	14.9	661	13	BQ743361
9	74	14.8	569	13	BQ789501
10	72.2	14.4	830	29	CG34022
11	72.2	14.4	876	29	CG310929
12	71.2	14.2	706	13	CA143501
13	71	14.2	598	10	BE426294
14	71	14.2	982	13	EX415111
15	70.4	14.1	668	12	BQ481484
16	70.2	14.0	563	29	CC679053
17	70.2	14.0	564	29	CC679056
18	69.8	13.9	744	14	CA270511
19	69.8	13.9	925	29	CNS0091P
20	69.6	13.9	616	13	BU100715
21	69.4	13.9	570	14	CA549499
22	69.2	13.8	1139	28	BZ552857
23	68.8	13.7	875	12	BG344702
24	68.8	13.7	935	29	CNS006XX
25	68.2	13.6	629	12	B1959998
26	67.2	13.4	476	10	BE215262
27	67.2	13.4	709	14	CD433365
28	67	13.4	935	29	CNS006XX
29	66.8	13.3	925	29	CNS0091P
30	66.8	13.3	982	13	EX415111
31	66.2	13.2	1315	11	AY106966
32	65	13.0	438	10	BE597495
33	63.8	12.7	602	13	BU998998
34	63.8	12.7	742	29	CG297240
35	63.8	12.7	1070	28	BZ557810
36	63.6	12.7	2079	11	AY108238
37	63.2	12.6	600	28	BZ895297
38	63.2	12.6	703	14	CA196875
39	63	12.6	642	13	CA025271
40	63	12.6	1101	29	CNS01751
41	62.8	12.5	726	10	BE519742
42	62.6	12.5	703	14	CA162305
43	62	12.4	643	13	CA162305
44	62	12.4	646	13	CA162262
45	62	12.4	932	29	CNS0072Q

## ALIGNMENTS

RESULT 1	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
LOCUS	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
DEFINITION	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
ACCESSION	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
VERSION	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
KEYWORDS	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
SOURCE	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
ORGANISM	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
REFERENCE	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
AUTHORS	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
TITLE	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
JOURNAL	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002
COMMENT	CA012845	HT06J23r	684 bp	linear	EST 23-OCT-2002

Barley ESTs from germinating seeds  
Contact: Stein Nils  
Unpublished (2002)  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert Length: 684 Std Error: 0.00  
Plate: 6 row: J column: 23

FEATURES  
source  
Seq primer: M13rev.  
Location/Qualifiers  
1..684  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="barke"  
/sub\_species="vulgare"  
/db\_xref="GABI:261368"  
/db\_xref="taxon:112509"  
/clone="HT06J23"  
/tissue\_type="endosperm early"  
/dev\_stage="0-16 hours after imbibition"  
/lab\_host="XL10-Gold"  
/clone\_lib="HT"  
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN  
Query Match 18.0%; Score 90; DB 13; Length 684;  
Best Local Similarity 52.3%; Pred. No. 0.0018;  
Matches 253; Conservative 0; Mismatches 220; Indels 11; Gaps 2;  
QY 6 CTACACGTCGGGACACCGGGCGGCCAAAGGCGGTCGTCCACGACGCGCACTGCCT 65  
DB 201 CTACTGTCGGGACACGCGGCTCCCAAGGCGTCATGCTACCCACCGCAGCCTCGT 260  
QY 66 CTGGTCCGTCG-----TTCCTGCTAGTGCCTTCCCGGGTTGTCGACACGAC 117  
DB 261 CACCAGCGTCGCCACGAGGTGACGCGGGAGAACCCGAACTGTACTTCGGGGAGGAGAC 320  
QY 118 CGGGTCTCTGGCGCGCTCCCGCTGTTCACAGCTTTCCGACATCGCTCGCTCTGTCC 177  
DB 321 GTGGTCTGTGGTGTGCGCTGTTCACATCTACTCGCTCACTCGGTGCTGTGGCG 380  
QY 178 GCCACGTCGTGGGGCCACGCGTCGCGATGCGCGACGCGAGCTCCGCCACGACGTGATG 237  
DB 381 GGCTCGGGCGCGGTCGCGATCGTATCATGCGCAAGTTCACACACGCGGTGCTGGTG 440  
QY 238 CGGCTGATCAGCGCAGAGCTGACCTTCCTGGCGGCGGTGCGGACCACTACACAC 297  
DB 441 CGCCTGTGGCGCGCACGCGGGTCACCGTGGCGCCCTTCGTGCCGCCATCGTCGTCGAG 500  
QY 298 CTGGTGGCGCGCCCGCGCAGCGCGTTCCTCGCGCGGAGCTGCGGATCGGCTCGGCC 357  
DB 501 ATGCCAAGACGACCGGGTACCGCGCGCGACCTCGCCCTCATCGCATGTCATGTCC 560  
QY 358 GGGGGCGCGGTCTCTGGCGCGCGGTCGCGAAGCGAGTTC---GAAGAGACCTTCGGGGTC 414  
DB 561 GCGCGCGCGCCATGCGCAAGGACCTCCAGGAGCGCTTCATGCGCAAGATCCGCCAAGCC 620  
QY 415 CCCTGATCAGCGCTACGCGACGACCGGACCTGCGGGCGGATCACCATGAACCCCGCG 474  
DB 621 GTCTCGGACAGGGGTACGGGATGACGAGGCGCGCGCGCTGTGTCTCATGTGCTCGCG 680  
QY 475 GACG 478  
DB 681 TTCG 684

RESULT 2  
CD865561  
LOCUS  
DEFINITION  
AZO2:101D12F001107 AZO2 Triticum aestivum cDNA clone AZO2101D12,  
mRNA sequence.  
ACCESSION  
CD865561  
VERSION  
CD865561.1 GI:32549377  
KEYWORDS  
EST.

SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
REFERENCE  
1 (bases 1 to 564)  
AUTHORS  
Genoplante.  
TITLE  
Genoplante, a major partnership french program in plant genomics  
JOURNAL  
Unpublished (2003)  
COMMENT  
Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).  
FEATURES  
source  
1..564  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZO2101D12"  
/tissue\_type="root"  
/clone\_lib="AZO2"

ORIGIN  
Query Match 16.9%; Score 84.8; DB 14; Length 564;  
Best Local Similarity 53.4%; Pred. No. 0.0094;  
Matches 251; Conservative 0; Mismatches 207; Indels 12; Gaps 3;  
QY 8 ACACGTCGGGACGACCGGGCGGCCCAAGGCGTGTGTCACGACGCGCACTGCCTCT 67  
DB 98 ACACCTCGGACGACGAGCCGCGCCAGGGGTGCGCTCAGCAGCGCACTGGCG 157  
QY 68 GGTCTGCTGCTTCTCTGCTACGTCGCGTTCCTCCCGGGTTGCGACGAGACCGGGTCTCT 127  
DB 158 CCTCGTCCAGAACATCCGCGCGGTGTACC---GGTTCCGCGAGACGCGACGCGTGG 214  
QY 128 GCGCGTCCCGCTGTTCACAGCTTTCCACATCGCTGCGTCTGTCGCGCACCGTGG 187  
DB 215 TGACACTGCGCTGTTCACGTCGCGCTCATGTGCGCGTGTCTCTCTGCGCT 274  
QY 188 TCGGGGCCACGCTCCGATCGCGACGCGCAGC-----TCGCGCAGACGTGATCGCGC 241  
DB 275 CCGCGCGTGGTGGCGCTCCCGCGCGCGCGCTTCCTCGGGGTCCAGTTCGCGCG 334  
QY 242 ---TGATCGAGCGGAGAGCTCGACCTTCCTGGCGCGCGTGGCGACCACTACCAACC 298  
DB 335 ACATGCTCGCGCGCGCGCCACGTGTACACGCGGTGCCACGATCCACCATCATCC 394  
QY 299 TGTGCGGGCGCGCGCGCGCGGTTCCTCCGCGCGAGCTTCGCGATCGGCTGGCG 358  
DB 395 TGGACCGCAGCGCTGCTGCGCGCGCGGTACCCGCGCTGCGGTTGTCGCGAGCT 454  
QY 359 GGGCGCGCTTCCTCGCGCGCGGCTCGAAGCGAGTTCGAAGAGACTTCGCGGGTCCCG 418  
DB 455 GCAGCGCGTCTGCGCGCGCGGTGATCTGAGCGCGCTGAGGCGGCTTCGCGGCGCG 514  
QY 419 TGATCGACGCTACGCGCAGCACCGAGACCTTCGGGGCGGATCACCATGAAC 468  
DB 515 TGCTGGAGGCGTACCCCATGACGAGAGCGCTCCCATCATGATGACCTCCAAC 564

RESULT 3  
BI958356  
LOCUS  
DEFINITION  
BI958356 666 bp mRNA linear EST 22-OCT-2001  
HVSME0014124f Hordeum vulgare rachis EST library HVCDA0015  
(normal) Hordeum vulgare subsp. vulgare cDNA clone HVSME0014124f,  
mRNA sequence.  
ACCESSION  
BI958356  
VERSION  
BI958356.1 GI:16309611



279 AGGACGTTGCTGCTGTGCTTCTGCGCTGTTCACATCTACTCGCTCAACTCGGTGCTGC 338  
 173 TGTCCGCACACGTTGTTGCGGGCCAGCGTCCGATCCGACACGGCAGCTCGGCCGACGAG 232  
 339 TGGCGGGGCTGTCGCGCGCGGTTCGGCGCATGTGTGATCATGCGCAAGTTTCGACATCGGCGCGC 398  
 233 TGATCGCGCTGATCGAGCGGAGAGCTCGACCTTCTGCGCGGCGTGGCGACACCACTTACC 292  
 399 TGGTGGAGCTTGTGTCGCGCGCGACACGGCATACCGTGGGCGCTTCGTGTCGCGCCCATCGTTGG 458  
 293 ACCACCTGTTGCGGGCGCGCGCGCAGCGCGCTTCTCCGCGCGGACCTTCGCGATCGGCC 352  
 459 TGGAGATGCGCAAGAGCCCCAGGTACACGCGCGGACCTCGCGTGCATCCGCGATGGTCA 518  
 353 TGGCGCGGGGCGCGGTCCTCGCGCGCGGGTTCGGAAGCGAGTTC---GAAGAGACCTTCG 409  
 519 TGTCCGCGCGCGCGCCCATGGCGCAAGGAGCTGCGAGGACGCTTTCATGGGCCAAGATCCCCA 578  
 410 GGCTCCCGCTGATCGACGCTACGGCAGCACCGACCTTCGCGGGCGATCACCATGAACC 469  
 579 ACGCCGTGCTCGGCGCAGGGGTTGCGGATGACTGAAGCCGCGCGGTCTGGCCATGTGCC 638  
 470 GCGCGG 475  
 639 TGGGG 644

CA092745 588 bp mRNA linear EST 23-SEP-2000  
 SCVPAM2068F01.g Am2 Saccharum officinarum cDNA clone SCVPAM2068F01  
 5', mRNA sequence.  
 CA092745  
 CA092745.1 GI:34946052  
 EST.  
 Saccharum officinarum  
 Saccharum officinarum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Saccharum.  
 1 (bases 1 to 588)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 Contact: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parruda@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
 http://www.bccccenter.fcav.unesp.br  
 Plate: 068 row: F column: 01  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers  
 1..588  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCVPAM2068F01"  
 /lab\_host="DH10B"  
 /clone\_lib="AM2"  
 /note="organ: Apical meristem and tissues surrounding of  
 immature plants; Vector: pSport1; Site\_1: Sali; Site\_2:  
 NotI; An unidirectional cDNA library generated from  
 [Apical meristem and tissues surrounding of immature  
 plants]. cDNA was prepared from poly+ mRNA using  
 SuperScript Plasmid System Kit (Invitrogen). The  
 double-strand cDNAs were fractionated in a sepharose  
 CU-2B 40cm-columns and fragments sizing between 0.8 and

FEATURES  
 source



JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605011 row: H column: 06.  
Location/Qualifiers  
1. .543  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site\_1: EcoRI;  
Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt  
lab"

FEATURES  
source  
1. .543  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site\_1: EcoRI;  
Site\_2: XhoI; Kernel endosperm cDNA library from Schmidt  
lab"

ORIGIN  
Query Match 16.0%; Score 80.4; DB 9; Length 543;  
Best Local Similarity 52.9%; Pred. No. 0.039;  
Matches 202; Conservative 0; Mismatches 171; Indels 9; Gaps 1;  
Qy 2 TGTCTACAGTCGGGACGACGCGGGCCCAAGGCGGTGTTCACGAGCGCAACT 61  
Db 155 TGCCCTACTCTGTCGGCAGCAGCGGGGTGCGCAAGGCGGTGATCTGACGACCGCAGCC 214  
Qy 62 GCCTCTGGTCGTGCTTC-----CTGCTACGTGCGGTTCCTCCGGGTTCGAGCC 112  
Db 215 TGAGCACAGCGTGGCGCAGCAGGCGGACGCGCAACCACTCGGCGTACCGCG 274  
Qy 113 AGAGCGGGGTCTGTCGCGGTCTCCGTGTTCACAGCTTTTCACATCGCTCGCTCC 172  
Db 275 CGAGCGTGGTCTGTGCTGCTCCCTCCATGTTCACATCTACTGCTACACCATCTGA 334  
Qy 173 TGTCCGCCACCGTGTGGGGCCAGCGTCCGGATCGCGACGCGAGCTTCGCCGACGACG 232  
Db 335 TGTGCGGCTGCGGTGGGGCCGCGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 394  
Qy 233 TGATCGGCTGATCGAGCGAGAGCTGACCTTCCTGGCGGGGTGCGGACGACCTACC 292  
Db 395 TGATGGAGCTGTGGAGCGCCACCGGATFACCGTGTGGCGCGCGTGTGGCGCGCGCATGTGG 454  
Qy 293 ACCACCTGGTGGCGCGCGCGCGCGCGGTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 352  
Db 455 TCGCGTGGCGCAAGACGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514  
Qy 353 TGGCGGGGGCGCGGTCTCTCG 374  
Db 515 TCTCGGGCG 536

RESULT 7  
LOCUS CD056352  
DEFINITION HO11P158 HO Hordeum vulgare cDNA clone HO11P15 5-PRIME, mRNA  
ACCESSION CD056352  
VERSION CD056352.1 GI:30599098  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 536)  
AUTHORS Zierold, U. and Schweizer, P.  
TITLE Barley ESTs from pathogen-attacked leaf epidermis

JOURNAL Unpublished (2003)  
COMMENT Contact: Patrick Schweizer  
Transcriptome Analysis, Cytogenetics Department  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, D-06466 Gatersleben, Germany  
Tel: 0049 (0)39482-5660  
Fax: 0049 (0)39482-5595  
Email: schweiz@ipk-gatersleben.de  
Insert Length: 536 Std Error: 0.00  
Plate: 11 row: P column: 15  
Seq primer: SK.  
Location/Qualifiers  
1. .536  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/cultivar="Ingrid BC mlo-5"  
/db\_xref="GABI:707315"  
/db\_xref="taxon:4513"  
/clone="HOLLIP15"  
/tissue\_type="leaf epidermis, 6 h and 24 h post  
inoculation with Blumeria graminis"  
/dev\_stage="7 d after germination"  
/lab\_host="XL10-Gold"  
/clone\_lib="HO"  
/notes="Vector: pBluescript SK+; Site\_1: EcoRI (5'-end of  
cDNA); Site\_2: XhoI (3'-end of cDNA); Approximately 5 % of  
the clones correspond to cDNA from the fungi B. graminis  
hordei and tritici, respectively. Due to a cloning  
artefact caused by the kit, in most cases the EcoRI site  
is NOT present, as well as the EcoRI adapter used for  
cloning. To excise the insert, restriction sites upstream  
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also  
due to the cloning system used Blue/white selection for  
recombinants is not 100% reliable. Average insert size is  
1.2 kb"

ORIGIN  
Query Match 15.0%; Score 75.4; DB 14; Length 536;  
Best Local Similarity 51.7%; Pred. No. 0.2;  
Matches 197; Conservative 0; Mismatches 181; Indels 3; Gaps 1;  
Qy 103 TTGTTCGACGACGAGACCGGTGTCTGCGCGCTTCGCGGTTCGACAGCTTTCCACATC 162  
Db 72 TTCCGGGAGGAGGACGT 131  
Qy 163 GCCTGCTCTGTCCGCCACCGTGTGGGGCCAGCGTCCGATCGCGCGCGCGAGCTCC 222  
Db 132 TCGTGTCTGTGGCGGGGTGCGGGCGCGGTGCGGATCGTGATCATGCGCAAGTTCGAC 191  
Qy 223 GCCGACGACGTGATCGCGCTGATCGAGCGGAGAGCTGACCTTCCTGGCGGGGTGCGG 282  
Db 192 CACGGGT 251  
Qy 283 ACCACCTACACACCTGT 342  
Db 252 CCCATGT 311  
Qy 343 CGGATCGCGCTGTGGCGGGGGCGGTCTCTCGCGCGCGGTGCGAAGCGAGTTC---GAA 399  
Db 312 CGCATGT 371  
Qy 400 GAGACCTTCGGGGTCCCGCTGATCGACGCTTACGCGAGCAGCAGACCTCGGGGGCGATC 459  
Db 372 AAGATCCCAACGCGGT 431  
Qy 460 ACCATGAACCGCGGACGCG 480  
Db 432 TCCATGTGCTGGCGTTCGCC 452

RESULT 8  
LOCUS BQ743361  
DEFINITION BQ743361 linear  
AUTHORS BQ743361  
TITLE EST 17-JUL-2002

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: [candersn@pw.usda.gov](mailto:candersn@pw.usda.gov)  
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: SK primer.

SOURCE ORGANISM	Triticum aestivum (bread wheat)
REFERENCE	Tukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum. 1 (bases 1 to 569)
AUTHORS	Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K., Dvorak, J., Lazo, G.R., Rausch, C.J., Wilson, C. and Woo, J.
TITLE	The structure and function of the expressed portion of the wheat genomes - Chinese Spring whole plant cDNA library
JOURNAL COMMENT	Unpublished (2002) Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: canders@pw.usda.gov Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: SK primer.

```

FEATURES
source
Location/Qualifiers
1. .569
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4162_B02_D04"
/tissue_type="Roots, leaves, crown, stem and sheath"
/dev_stage="Adult"
/lab_host="E. coli SOLR"
/clone_lib="Wheat CS whole plant cDNA library"
/note=vector: lambda Uni-ZAP XR, excised phagemid
pbLuescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant
tissues from wheat cv. CS grown to full tillering stage in
greenhouse were collected at University of California,
Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
RNA was prepared from leaves (young leaf and third leaf),
whole roots, crown, stem and sheath tissues, and then
equal quantities of RNA were pooled from these
samples. PolyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo excised
to give phuescript SK(-) phagemids in J. Dvorak's lab (E
Akhunov, J. Dvorak) at the University of California, DNA
Davis. Colony plating, plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

```

ORIGIN

Query Match 14.8%; Score 74; DB 13; Length 569;  
 Best Local Similarity 51.6%; Pred. No. 0.32;  
 Matches 224; Conservative 0; Mismatches 200; Indels 10; Gaps 2;

QY 2 TGTCTTACAGCTCGGGCACCACCGGGCGGCCAAGGCGCTGCTGTCTCCACGACGCAACT 61  
 Db 128 TGGCTTACTCTCTCGGCACACACCGGGCTGCCAAGGGCGTCATGCTCACGACGCGACGC 187

QY 62 GCCTCTGTGTCTCGCTGTC-----CTGCTACGTGCGCTTCCCGGGTGTTCGAC 112  
 Db 188 TCATCACAGCGTCCGCCAGCAGGTGACCGGGGAGAACCCGAACTGTACTTCAGCAAG 247

QY 113 AGGACCGGGTGTCTGCGCGCTCCCGCTGTTCACAGCCTTTTCACATCGCTTGTCTCC 172  
 Db 248 AGGACGTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307

QY 173 TGTCCGCCACCGTGTGTCGGGGCCAGCGCTCCGGATCGCGCGAGCTTCGCGCCGACGACG 232  
 Db 308 TGGGGGCTTGGCGCGCGCGGTGCGGATGCTGATCATGCGCAAGTTCGACATCGCGCGCG 367

QY 233 TGATCGCGCTGATCGAGCGGAGAGCTCGACTTCTGCGCGCGGTGCGCGGACCTTAC 292  
 Db 368 TGGTGGAGCTGCTGCGCGCGCACGCGCATCAGATCGCGCGCTTCTGCTG-CGCGCATCGTG 426

QY 293 ACCACTGTGTCGGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 352  
 Db 427 TGGAGATCGCGCAAGAGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 486

QY 353 TGGCGCGCGCGCGCGCTTCTCGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 412  
 Db 487 TGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 546

QY 413 TCCCGCTGATCGAC 426  
 Db 547 ACCGCGTGTCTGCG 560

RESULT 10  
 CC634022  
 LOCUS OGULM87TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBma0468P05, linear GSS 19-JUN-2003  
 DEFINITION genomic survey sequence.  
 ACCESSION CC634022  
 VERSION CC634022.1 GI:32010601  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 830)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 TITLE OGULM87TH  
 JOURNAL Other GSSs: OGULM87TH  
 COMMENT Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.  
 Location/Qualifiers  
 1..830  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"

/clone="ZMMBma0468P05"

/clone lib="ZM\_0.7\_1.5\_KB"  
 /note="Vector: pBCSK-; Site:1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Query Match 14.4%; Score 72.2; DB 29; Length 830;  
 Best Local Similarity 51.1%; Pred. No. 0.59;  
 Matches 201; Conservative 0; Mismatches 183; Indels 9; Gaps 1;

QY 2 TGTCTTACAGCTCGGGCACCACCGGGCGGCCAAGGCGCTGCTGTCTCCACGACGCG----- 56  
 Db 324 TGGCTTACTCTCTCGGCACACACCGGGCTGCCAAGGGCGTGTCTCTCGCACCGTGGGC 383

QY 57 ----CAACTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 112  
 Db 384 TGGTACACAGCGTGGCGAGCTGTTCGACGCGGACACCGGAACTTCTACTTCCGGAGG 443

QY 113 AGGACCGGGTGTCTGCGCGCTCCCGCTGTTCACAGCCTTTTCACATCGCTTGTCTCCGCTCC 172  
 Db 444 ACCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503

QY 173 TGTCCGCCACCGTGTGTCGGGGCCAGCGCTCCGGATCGCGCGAGCTTCGCGCCGACGACG 232  
 Db 504 TGTCCGGGATGCGCGCGCGCGCGCGCGCTTCTGATCATGACCGCTTCGACACTCTCCGCA 563

QY 233 TGATCGCGCTGATCGAGCGGAGAGCTCGACTTCTGCGCGCGGTGCGCGGACCTTAC 292  
 Db 564 TGTTCGAGCTGCTCAAGAGGCAACGCGCATCAGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 623

QY 293 ACCACTGTGTCGGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGCGCGCGAGCTTCGGGG 352  
 Db 624 TGGAGATGCTCAAGAGGCGACCGCATTCGACCGCGCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTG 683

QY 353 TGGCGCGCGCGCGCGCTTCTCGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 385  
 Db 684 TCTCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTCGCGCGCGAGCTTCGAGAGACCTTCGGGG 716

RESULT 11  
 CC310929/c  
 LOCUS OG3DO25TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBma0783E01, linear GSS 26-AUG-2003  
 DEFINITION genomic survey sequence.  
 ACCESSION CC310929  
 VERSION CC310929.1 GI:34228089  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 876)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 TITLE OG3DO25TV  
 JOURNAL Other GSSs: OG3DO25TV  
 COMMENT Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.  
 Location/Qualifiers  
 1..876  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"

FEATURES  
 source









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tgcgttcctggtcgctgcgt.....cgcgtggtcggtcggtg 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_nam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
4	435.4	86.9	3900	1	AF262754	AF262754 Amycolato
5	334.6	66.8	5676	6	A69718	A69718 Sequence 1
6	332	66.3	123580	1	AF263912	AF263912 Streptomy
7	332	66.3	125401	6	AX211739	AX211739 Sequence
8	328.2	65.5	20394	1	SNA13222	AJ132222 Streptomy
9	328.2	65.5	20394	6	AX067996	AX067996 Sequence
10	328.2	65.5	84985	1	SNA278573	AJ278573 Streptomy
11	328.2	65.5	104326	1	AB070940	AB070940 Streptomy
12	328.2	65.5	320150	1	AP005033	AP005033 Streptomy
13	325	64.9	82746	1	AF453501	AF453501 Actinosyn
14	320.8	64.0	113193	1	AF357202	AF357202 Streptomy
15	320.8	64.0	113193	6	AX703543	AX703543 Sequence
16	317.6	63.4	65140	6	AX211705	AX211705 Sequence
17	317	63.3	78210	1	AB070949	AB070949 Streptomy
18	317	63.3	300425	1	AP005022	AP005022 Streptomy
19	316	63.1	138203	1	AY310323	AY310323 Streptomy
20	314.4	62.8	77534	1	AF235504	AF235504 Streptomy
21	314.4	62.8	77536	6	BD235937	BD235937 Polyketid
22	312.2	62.3	107379	1	SHGCCP1R	AX271638 Sequence
23	312.2	62.3	290850	1	SC0939127	XS6780 S.hygrosco
24	311.8	62.2	39314	1	SGR300302	AJ939127 Streptomy
25	311.2	62.1	132544	1	AF521085	AJ300302 Streptomy
26	311.2	62.1	132544	1	AF521085	AF521085 Streptomy
27	309.6	61.8	28732	1	AF082100	AF082100 Streptomy
28	309	61.7	107379	1	SHGCCP1R	XS6780 S.hygrosco
29	309	61.7	281450	1	AP005032	AP005032 Streptomy
30	307.4	61.4	15872	6	AR362090	AR362090 Sequence
31	307.4	61.4	15872	6	AR362091	AR362091 Sequence
32	307.4	61.4	15872	6	BD217171	BD217171 DNA encod
33	307.4	61.4	30000	6	AX250261	AX250261 Sequence
34	307.4	61.4	43280	1	SFU78289	U78289 Streptomyce
35	307.4	61.4	103450	1	AF440781	AF440781 Streptomy
36	305.8	61.0	38506	6	BD232534	BD232534 Recombina
37	305.8	61.0	38506	6	AR271680	AR271680 Sequence
38	305.8	61.0	38506	6	AR277664	AR277664 Sequence
39	304.2	60.7	4674	6	BD235956	BD235956 Polyketid
40	304.2	60.7	4674	6	AR271657	AR271657 Sequence
41	304.2	60.7	4725	6	BD235955	BD235955 Polyketid
42	304.2	60.7	4725	6	AR271656	AR271656 Sequence
43	304.2	60.7	4737	6	BD235958	BD235958 Polyketid
44	304.2	60.7	4737	6	AR271659	AR271659 Sequence
45	304.2	60.7	4767	6	BD235957	BD235957 Polyketid

# ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS  
DEFINITION  
53784 bp DNA linear BCT 09-FEB-1998  
Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.  
ACCESSION  
AJ223012  
VERSION  
AJ223012.1 GI:2764760  
KEYWORDS  
ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
SOURCE  
Amycolatopsis mediterranei  
ORGANISM  
Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.  
REFERENCE  
1  
AUTHORS  
Schupp,T., Toupet,C., Engel,N. and Goff,S.  
TITLE  
Cloning and sequence analysis of the putative rifamycin polyketide

Pred. No. is the number of results predicted by chance to have a

synthase gene cluster from Amycolatopsis mediterranei  
 Unpublished  
 2 (bases 1 to 53784)  
 Schupp, T.  
 Direct Submission  
 Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
 AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
 SWITZERLAND

## FEATURES

source

1..53784

/organism="Amycolatopsis mediterranei"

/mol\_type="genomic DNA"

/strain="IBG A3136"

/db\_xref="taxon:33910"

1336..15543

/function="CoA ligase, acyltransferase, ketoreductase"

b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"

/note="ORF1"

PKS modules 1-3"

/codon\_start=1

/transl\_table=11

/product="rifamycin polyketide synthase, type 1"

/protein\_id="CAA11036.1"

/db\_xref="GI:2764761"

/db\_xref="GOA:O54666"

/db\_xref="SPTREMBL:O52789"

/translation="MRTDLIKPLHVALLENATREFAKPAFAADHRTVTYGLLEARTTR  
 LAGHLAAGLVGRHGRDRAICLNRVSTVESYFALLRAGAVGVPLNPGSATAELHEPLTD  
 SGATVVTDAAGARLRLAPHVELLVTGDDVPGEHAGYDELALSEFAEPADLDELDE  
 PAMVYTSSTTRPGKGVSTQRNCLWSVASCYVFPFGLSDQDRVLPFLFSLSHIA  
 CWLSTVVGASVRIADSGSADDMRLIEASTFLAGVPTTVHLVRAARQRFSGAP  
 FLVLAGGAVLGALESEFEETVPLIDAYGSTCTCGAI TWNPDDARVGECSGLAV  
 PGVDVVPDPTGLDVPAGEGEVGVGNVLMGVHNSPEATAAMDRCWPTGDLAR  
 RDAGYFTICRIKELIIRGKANIHPGEVEAVLRTVDGVADAAGVGVDHTIGENPVA  
 YVIGPTGDPDAALIEKREQLSAKYVPDRIIEVAHPRTAGSKRRLGLTDRPQALR  
 YAAETHEEGSRHDESAALRARLSGLDERAQCELLEDLVRTQADVLGQVPPGGR  
 FDLGFTSLAIVELNRRLTEHTGLMLPASAVFDHTPTPAALAAARVAELGLIQVAEP  
 VVAADGPEPIATVGMACRLPGGVASPEDLWRLVAERVDVAVSEFPGRGMDLSDIDP  
 RERAGTSYVGGSGFLHDAGEFDAGFEGISPREAVMDPQORLLLETSEALENAGVDP  
 IALKHDTGVGSLMGQGGSGNAPELGGFTVTGVASSVGRSVYLGLGEPATV  
 DTACSSLVAMHLAQAALGQECMALAGVTVMATPGSFVSFRQALADGRCARF  
 AAAADGTGSEGVVLERLSVAREGRHILAVLRGSANODGASNGLTAPNGLSKQ  
 RVTRRALAAGLAPSDVDVVEAHGTGTGLDPIEAQALLATYGOERKQPLMLGSLNS  
 IGHQAAGAVGVKMWQLRHETLPTLVDKPTLEVDWSAGAILLELTAARPNRG  
 RRRAGSVSGFTGNTNHLILLEAPAEPEVPAPELVPVLVVSARSTESLSQAERLA  
 SILGDSVLTEVAGALVSRRAVLDRAVAVVAGSREAVTGLRALNTAGSGTGGKVVV  
 FPQGTQWAGMGRELLAESFVABRIAECAALAPWDLSLVLRGEGDLGRVDVQ  
 PACFAMVGLAAVSESVGRVPDPAVVGHSQGEITAAACVSGALSLEDAKVALRSQAIA  
 AELSGRGMSVALGEDDVSRVLDGVEVAAVNGSPSVVAGDAHALDALTLELSGEG  
 IRVRVAVYASHTRHVEDIRDTLAETLAGISAQAPVPFYSVTSEWVRDAGVLGG  
 RLLASMGELFVRGIDVDMTAVPAGVNDLPTIYAEHHRWLEPAEPASAGDPLLT  
 VVSTPGSDRLTAQVDMRAQPMVAGLVNLAALVEAAILRLGLAGTPVVGELVDAP  
 VLPRGSEVQLIIVGEGEQRRPTEVFSREADEPWHHTAGTLTAPAAVPEPAA  
 GDAVTYVAGLRDGHGHPALLDAARTVVGDDLLPSTGVSLASGATATVTP  
 TLTATLTPAGQVLTVESVGTPTFVAEQGTLDALFRVMDWPEPLPTATADFLPYE  
 ATSAREATLSALQALADPAETRLAVTGDCTEPPGAAIIMGLVRSASOEHPRGLVLAD  
 DPAVLPVAVSAGEPQVRVNGVAVPRLTRVTPRODAPLDEGTVLTIGTGTGLGA  
 LTAHLVTHAGVHLVLSRSGEAPLQBELTALGASVATAACVADRAQLEAVRLAI  
 PAEPHTPATVHTAGLVLDGVTTLTPDLATVRRPKVDAARLDELTDREADLAALFIF  
 SSAAGVLPNGQAGYAAANALDALARQNSLDLPASTAWGWTAVSGVTEHLGDAD  
 LRNRQIRGMSGLPADENMLDAAIATGTTVAAKTVAALRATAKAGVPPCLLRL  
 APIPRAAAKTASITRELACLAETEQAALLDLVRHAALVLSHGSVHSQRTFKD  
 AGFDSLTAVELNRLLAATGLTSLPAMIFDYPPADALDHLRAKLIGSAAARPAEIGT  
 AAAEEPIAI VAMARFPFGVSPEDLWRLVADGADVATEPDRGMDTDLRKHEDPDH  
 EGTTYVHGFADDAAGDAFPFGISFNEALANDPQORLLLETSLWSEFLFERAATDPTTL  
 AQDQIYGVAGFNGDYSMRHRAAGVEGRLTGCSSVLSGSVAVYHFGVEGPATVDT  
 ACSSLSVLHMAVQALQRCMALAGVVMWGTFTTFVFSRQRLADPGRCAPAD  
 GAGGTGSEGVGLHVERLSAQRGHQVLAUVRGSAVNSDGSANGLTAPNGSQORV  
 ITRKALAAAGLSTSDVDAEHAGTGTGLDPIEABALLATYQNRPTPLWLSVKSNLG  
 HTQAAGAVAGIKNVMWRHGVLPRLTHVDRPSSYVDWSAGAVELLTEARDWVSNRHP  
 RRAGSVSGTGGTNAHVLEVAAPITPQPEPAEFLVPLVVSARTAGTAAGQAGRLA  
 AFLGDRTDVRVPDAAVALATTRAQLOHRAVVLASDRQAQLCADLAAFGSGVVTGTVDG

KLAVLFTGQSQWAGMGRLEAETFPVFRDAFEAAACEAVDTHLRERPLEWVFDGALL  
 DOTVYTOGALPAVETALPRLPESWGRVPCGLLAGHSIGELAAAHVSGVLDDADAGLIVA  
 ARGRLMOALPAGGAMVAQVATEDEVAPLDGTVCVAAVNGPDSVVLSTGEAAVLAVAL  
 ELGRGKTRRLAYSHAFSHPLMRMLDDFRAVAFELTVRAGLSVSTVLTGELAAALD  
 SPDVWVGQVRNARVRSFADVATAGAGQASTFLELGGGALAAWALGTGLGPEQSCVATL  
 RKNAGVSPDLVLTALAEHLVHVGVDVMTVTLDEPATAVGTVLTPTFYAFQHFVWDVDET  
 AAVSTPTPAPPIVDRPQVDVLELVRESAAVVLGHRDAGSFDLRSFKDGHGDSLSAV  
 KLNRRLDFTGVLPSTLIFDYPNPVLAHLRAELLGERPAAPAFVDRDYDSDEPIAI  
 VGMSTRLPGGADSPEELWKVAEGDRDASVGFVDRGMDLGLYHPDPHAGTSYRS  
 GLHDAQOPDAGLFGISPREALANDPQORLLLETSEALERAGVDPLSARGSDGVFT  
 GLVHHDYTRLREVDPDQVMTGTASTASVASGVAYVYFPGEPATVDTACSSSIVA  
 MHLAAQALQEGCSMALAGATVMASPDALFESQRGLSDAGRCARCAEAGADGTGWA  
 EGVGVVLERLSVAREGRHILAVLRGSANODGASNGLTAPNGSQORVIRKALASA  
 GLASDVVDVVEHGTGTGLDPIEYQALLATYQEREQPLWLSKLSNLGHTQAAGV  
 VGVIKMIMWRHGVMPATLHVDRBTSQVDSWAGAILVLEAREWPRTRPRAGVSSF  
 GASCTNAHLIIEGPAEAEVDEAVSVVPLVVSARSAGSLAQAGLAALVLENSLAG  
 VAGALVSRGATLNERAVIAGSDEAQQGLQALARGENAPGVVTTAGPKGVVVF  
 QGSGQMGMDRLDSSPFFAARIKECAALAEQMTDMSLLDVGADDLDDVLRDVTQOP  
 ASFAMVGLAAVMTSLGTVDAVLGHSGEITAAACVSGALSLEDAKVALRSQAIA  
 ELAGRGMSVALSEEDAVARLTWNAVEVAAVNSPSSVVIAGDAQALDEALEALAG  
 DGVVRVAVDYASHTRHVEAIAETLAKTLAGIDARVPAIPYFTVLTGWTIQAUVDA  
 GYWRNLQQRVPGPSVADLAGLHTVFEIISAHPVLVQPLSEISDDAVVTGSLRDD  
 GGLRRLASAEELVYRGVAVAVDMTAAVPAAGWDLVETIYAFDRHRFWHEAREAEAGM  
 DGEFWTAIEQSDVDSLLELLELVEPQRGALSTVVPVLAQMRDRRRERSTAEKRTQVT  
 WQLEREAAGVPGGRWLA VWPAGTDTALLKELTQGLDITVLEIEFEASRAQLAEQRLN  
 VLAEHDLTGVLISLALDGPDAAEITASTLALVQALGDTTTSAPLWLTGSAVNI  
 QADVAPAAQAVNGLGRAVALERLDRWGLVDLPAAIDARTAQALLGVINGAAGEQL  
 VARSQVVRRLVRKVPESATSRWEPGTVLVTGAEGLGRHASWLAQSGAERLIV  
 TGTGDELTABLAEEFTTVECATDRDAIAQLVADSEVTAVHAADTAOTSSVDVT  
 LVADDEVFAAKVTAVMDLQFEDTDFADVFSIAGIWWGGGGQGGPAGANAVLDA  
 IWEVRRAKATKATSTANGALDQIGIMDEALAAQLRRRGVTPMAPLAVTAMVAVAG  
 NEKAVADMDWAAFTPAFTSPSPFLADPEAKAILRAADDGDEGDTASLADSL  
 RAVPDANRILLKLVGRHASTVLSHGAEGIGPQAQOEFGVDSIAAVNLNSLHAA  
 TGLRLPATLIDYPTPEALGVYRVELLEADGLDREDDLRLVLAAPFPAEFKAEAG  
 VLDTLGLADTGTGTGTAEETTAAPADDAELIALDISGLVQALGQTS"

CDS

15550..30759

/function="b-ketoacyl-ACP-synthase, acyltransferase, acyltransferase, ketoreductase, acylcarrier protein"

/note="ORF2"

PKS modules 4-6"

/codon\_start=1

/transl\_table=11

/product="rifamycin polyketide synthase, type 1"

/protein\_id="CAA11036.1"

/db\_xref="GI:2764762"

/db\_xref="GOA:O52789"

/db\_xref="SPTREMBL:O52789"

/translation="MANQSRKMSAPNEQIVDALRSLKENVRLOQENSALAAAAE  
 VPAIVSMACRYAGGIPEDFWRVSEGDVYTFPDRGVDVGLYHDPDPNPGTYI  
 VREGAFDAAQFADGFFGISPREALANDPQORLLLEVSWETLERAGIDPHSVRGS  
 GYVAGVHODYAPDLGSGPMSLERALGTAGVSGRVAYTLGLEPAPVVTAAACGS  
 SILVAIHLAAQALRGCSMALAGSTVMAATPGCFVGFARQALAFDGRCKSVAAADG  
 SWAEGVGVLLERLSVAREGRHIVLVRGSANODGASNGLTAPNGAQOQVIRKA  
 LASAGTVDSDVTVGEGHTVGLDPIEYQALLATYQEREQPLWLSKLSNLGHTQAAGV  
 QAGVAGVYKMWGLRQVLPATVTEAPVTPVPLVVSAAATASLAPGALLSGSGVP  
 GDISLADAGTILVSRGMLSERAVVAGVSHHEAVTLRALARGESAPGILSGSGV  
 GKVVVFPQGTQWAGMGBELDSSSEVFAARIAECETALGRVMDVSLTDLVEADLL  
 DRVDVQPSAFLVAGLAWASLSEVPEAVVHGSQGEITAAACVSGALSLEDAKVA  
 LRQAISLASLRGMSVALSEEDATLEFVPAVWAGVEVAAVNGPSSVVIAGDAEALD  
 EALDAGLDGQVYRNLQRVRFPAVAAELIEQGRVFEVVSAPVLPVINFELVDVTE  
 VEDAGVLDGQVYRNLQRVRFPAVAAELIEQGRVFEVVSAPVLPVINFELVDVTE  
 AVVTGTLREDDGRLRLASAEELVYRGVTVDMGSLVLPSSRVELPTYAFDQHWLQ  
 MGSAADAVSLGAGADHLGVNVLPOSGLVFTSRLSLKSHPLMAGHAGVVL  
 PGTVYDIALRAGDELFGVLELIEALPLVIGRGGVYVQVAVSPNITGRGVAVF  
 SMREDEWTRHATGLLGASTSRERSPRFAAPFAGAPPIDVENEITDLTERGAYIS  
 QFQGRMVRVRRGDEVEFVALPDHREDAGFGLHPALLDAALDAALNANPDDRSV  
 LFPANGLVHLHAGVSAVLRVAPGPDALTFAQDEATQGLVVTMDLSVRESVSAQL  
 ETAAGERLDFQVDMVIEPATTAETAHAELVLEAFGEAPLELTSRLVLEAVGSMALD  
 ADAEARLVVVTGAPREVTPAGAVMGLVRAQAENPGRIIIVDTGDVPLGAVARH  
 GEPFLVRGNARVFRPLAATGEPAPVAFPEGTVLTGTGSGLGVAGLHVARH  
 GYRRLVARSRGVAAEDVTEIETEQAQVSVVACDVQHLDELTRDLGLDAFVVFSSAALMGSAQGG  
 GLIUDGVI GALNREELAGVAPKVDVQHLDELTRDLGLDAFVVFSSAALMGSAQGG

NYAANAFLDGLMAGRAAGLPGVSLAWLWEADGLTANLSATDOARMSRGVLPMT  
PAEALDIPDGLAEQALLVPILKDLRLTQRQATAGVEVPHLLRGLVAPASRVTRTA  
ASGGGLVHKAGAPAEQAVLLGIVQAEAAVLGFFNAPELAQTRGFSDLGDSLT  
AVELNRNLASAATGKLPAILVDFYPTVALARHLRBEIGETVAGAPATPVTTFADAGE  
PIAIVGWACRLPGGVMSPKVRWVABRGDPPFDRGDWDLGLDFPPTFATAY  
IRQGFSLHAAALFDPFGFISPREALAMDPOORLILLEASWEALRAGIDPTKARGDVA  
GVSGVSTHLYLESLSNMPALEGFTVTTATAGSVASGRVSTFGFEGPATVDTACSS  
SLVALHLAAQRLOCECTMALAGVAVMGSPGVIGTMSRORGMADDEGVKAPADGAG  
TVLSEGVIVLIERLSVAREGRHVLAVLRGSVNVQDASNGLTAPNGPSQORVRS  
LAGAGLQSRVDVVEAHGTGALGEPTEAQLLATYIGKSRFPFLWGLSKSNIHGTQA  
AAGVAATVMQALRQDTPPTLPTLHQEPTKQVDSAGAVELLTGREWRNGHPRRAG  
VSSFGISGTMNHLILEAPADTAEADVPDAPVVPVVISARSTGSLAGQAGLAALFDG  
DVPLTRVAGALLTTRATLTVRAVVGASAEARAGLTALARGESGLVGTAGMPGK  
TMVFPFGCTWAGGRRELLEASVPEAEIECECAAOLPMIDMSLVLVREGELDRV  
DVLQACFANVGLAAVWASGVVDPDAVLHSGQIEIAAACVSGALSLEDAKAVVALRS  
QALAEISGSGMASIQLSHDEVAARLWAPAGRVEIAAVNPGFVSIVAGDAEALTEAV  
EVLGRRVADYAGTHRVEDIQTDLAETLAGIDAQAPVFPVSTVAGEWITDAGVVD  
GGVYRNLRNOVGFPAAVELIEQHGVFVVSAPHLVQPISELTDAVVTGTLRRDD  
GGVRLTSMALFVRGVPDWTAPPARVELPTVAFDHOHFWLSPPAVADAPALGH  
AGADHPLGLVLPQSDGLVPTSLSVRTHPLADGVPAALVELAVRAGDEAGCPV  
LADITVEKLLVLPESGGURVQVIVSGERTVYVSOLEGAEDWIRNATGHLATAPAE  
AFDTWAPPAQGVQDGLWRGDEIFAEVALPEELDAGAFGLHPLDAAVQVPLADD  
EQPAEWSLVLHAGASALRVLPVGGALQAADEGTGLVLTADSVAGRELASGKTRAG  
SLYRVDETVSTADASAVANIEVEAFGEPELTGRLVEAVQTLVLTAAADARLVV  
TRGARVETDPAGAAVWGLVRAAOAENPGRIFLIDTGEIPLATGDEPIAIVRGKFF  
VPLTRAPSGAAVPRPDGTVLISGAGALGGLVARELVERHGVKLVLASRRGRDAG  
VADVLADLAADVSVVACVDSBRAQVALLDHERPTAVVHTAGVIDAGVETLDRRLA  
TVFAPKVDARHLDLDRDDDAFEVYSSVAFMGAGSGSYAANAPFLGLMANRR  
AAGLPLGLSLANDQSTGMAAGTDEATFARMRSRGGIIMTQAEGMDFDRAUSSAE  
SILVPAKDLDRGADAAAGVPHMLRGLVRAGRAQAAASTVDNGLAGRLAGLAPA  
DQTLTLLDLVRAQVAAADSAVRVDTAPKADGFDLSLTAVELNRNMTATGLKLP  
ATLVDPYNPQALASHLDELGGAACTPVTTAAAKADLDEPIAIVGMACRPLPGVAGP  
EDLRLVAGDVAVSSPFTDRGWDHDSLYDDPPARPCKTYTRHGCFLEAGLFDAGFF  
GLISPREAVMPDQQLLEASWEAMEDASLSLKGNDDVGTGFGGVVAPGDSV  
VPLEGFAGTGSSSVASGRVSYVFFGEPAPVTTIDSACSSLVAMHLAAVGVLLERLSV  
SMALAGGATWMANPGAFVFSRQGLAVDGRCKAFAPAAADGTGMAEGVGVLLERLSV  
ARERHRLILAVLGAANNQDASNGLTAPNGPSQORVIRRALVSLAGLAVDMDVVEAH  
GTGTLTGPILAEQALLVTKDRESPLWGLSKSNIHQAAAGVAGVIMVQOALRHE  
VLPLTLVDRPTPEVDMNAGAVELLTEAREWFRNRPBRAGVSAPFVSGSTNAHLILEE  
APAEPEPPTPEVPLPVVVSASRLASQOAGRLAGFVAGDASLAGVAKALVNRKRAL  
TERAVMVVGSREAEVTNLEALARGEDPAAVTTRAGSPGKLVVFPFGQSGSVWIMGR

Query Match 100.0%; Score 501; DB 1; Length 53784;  
Best Local Similarity 100.0%; Pred. No. 6.5e-45;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGTCTTCTGCTGCTGCATGCGCGTGCAGGCGCTCGAGCGCGCGAGTGTCTCC 60  
DB 8434 TCGTCTTCTGCTGCTGCATGCGCGTGCAGGCGCTCGAGCGCGCGAGTGTCTCC 8493

QY 61 ATGGCGCTCGCGGGCGGCTGATGATGGGCAAGGCGTTCGCGAGCGCGCGAGCGC 120  
DB 8494 ATGGCGCTCGCGGGCGGCTGATGATGGGCAAGGCGTTCGCGAGCGCGCGAGCGC 8553

QY 121 CGGCGACGCGGGCTGGCCCCCGAGCGCGCTCGCGAGCGCGCGAGCGC 180  
DB 8554 CGGCGACGCGGGCTGGCCCCCGAGCGCGCTCGCGAGCGCGCGAGCGC 8613

QY 181 ACCGGCTGCTCGAGGGCGTCCGGGCTGCTCTGCTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 8614 ACCGGCTGCTCGAGGGCGTCCGGGCTGCTCTGCTGGAGCGGCTGTCCGAGGCTCAGCGT 8673

QY 241 CGCGGGCACAGGCTCTCGCGCTGCTGCGGGTTCGCGGGTCAACTCCGAGCGCGGCTCG 300  
DB 8674 CGCGGGCACAGGCTCTCGCGCTGCTGCGGGTTCGCGGGTCAACTCCGAGCGCGGCTCG 8733

QY 301 AAGCGCTTACAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
DB 8734 AAGCGCTTACAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8793

QY 361 GCCCGCGGACTGTTCACATCGGACGCTCGAGCGCGCTGGAGCGCGCACCGGACCGAGC 420  
DB 8794 GCCCGCGGACTGTTCACATCGGACGCTCGAGCGCGCTGGAGCGCGCACCGGACCGAGC 8853

QY 421 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 480  
DB 8854 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 8913

QY 481 CGCTGTGCTCGGGTCTGGGTG 501  
DB 8914 CGCTGTGCTCGGGTCTGGGTG 8934

QY 421 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 480  
DB 8854 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 8913

QY 481 CGCTGTGCTCGGGTCTGGGTG 501  
DB 8914 CGCTGTGCTCGGGTCTGGGTG 8934

RESULT 2  
A69720 53789 bp DNA linear PAT 07-MAY-1999  
LOCUS Sequence 3 from Patent WO9807868.  
ACCESSION A69720  
VERSION A69720.1 GI:4774328  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 53789)  
AUTHORS Schupp,T., Toupet,C. and Engel,N.  
TITLE RIFAMYCIN BIOSYNTHESIS GENE CLUSTER  
JOURNAL Patent: WO 9807868-A 3 26-FEB-1998;  
CIBA GEIGY AG (CH)  
FEATURES  
source 1. 53789  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

ORIGIN  
Query Match 100.0%; Score 501; DB 6; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 6.5e-45;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGTCTTCTGCTGCTGCATGCGCGTGCAGGCGCTTCGAGCGCGCGAGTGTCTCC 60  
DB 8434 TCGTCTTCTGCTGCTGCATGCGCGTGCAGGCGCTTCGAGCGCGCGAGTGTCTCC 8493

QY 61 ATGGCGCTCGCGGGCGGCTGATGATGGGCAAGGCGTTCGCGAGCGCGCGAGCGC 120  
DB 8494 ATGGCGCTCGCGGGCGGCTGATGATGGGCAAGGCGTTCGCGAGCGCGCGAGCGC 8553

QY 121 CGGCGACGCGGGCTGGCCCCCGAGCGCGCTCGCGAGCGCGCGAGCGC 180  
DB 8554 CGGCGACGCGGGCTGGCCCCCGAGCGCGCTCGCGAGCGCGCGAGCGC 8613

QY 181 ACCGGCTGCTCGAGGGCGTCCGGGCTGCTCTGCTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
DB 8614 ACCGGCTGCTCGAGGGCGTCCGGGCTGCTCTGCTGGAGCGGCTGTCCGAGGCTCAGCGT 8673

QY 241 CGCGGGCACAGGCTCTCGCGCTGCTGCGGGTTCGCGGGTCAACTCCGAGCGCGGCTCG 300  
DB 8674 CGCGGGCACAGGCTCTCGCGCTGCTGCGGGTTCGCGGGTCAACTCCGAGCGCGGCTCG 8733

QY 301 AAGCGCTTACAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
DB 8734 AAGCGCTTACAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 8793

QY 361 GCCCGCGGACTGTTCACATCGGACGCTCGAGCGCGCTGGAGCGCGCACCGGACCGAGC 420  
DB 8794 GCCCGCGGACTGTTCACATCGGACGCTCGAGCGCGCTGGAGCGCGCACCGGACCGAGC 8853

QY 421 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 480  
DB 8854 CTGGCGCACCGGATCGAGGCGCGAGCGCTGTGCGGCACCTACGCGCAGAACCGGGAACG 8913

QY 481 CGCTGTGCTCGGGTCTGGGTG 501  
DB 8914 CGCTGTGCTCGGGTCTGGGTG 8934

```

RESULT 3
AF040570
LOCUS AF040570 90445 bp DNA linear BCT 16-APR-2001
DEFINITION Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
ACCESSION AF040570 AF040571
VERSION AF040570.2 GI:123331604
KEYWORDS
SOURCE
ORGANISM Amycolatopsis mediterranei
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
REFERENCE
1 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R., Yu, T.W.,
Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and
Floss, H.G.
TITLE Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
the molecular analysis of the rif biosynthetic gene cluster of
Amycolatopsis mediterranei S699
JOURNAL Chem. Biol. 5 (2), 69-79 (1998)
MEDLINE 98174059
FUBMED 9512878
REFERENCE
2 (bases 1 to 90445)
Kim, C.G., Yu, T.W., Fryhle, C.B., Handa, S. and Floss, H.G.
TITLE 3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
formation of the precursor of mC7N units in rifamycin and related
antibiotics
J. Biol. Chem. 273 (11), 6030-6040 (1998)
MEDLINE 98165773
FUBMED 9497318
REFERENCE
3 (bases 1 to 90445)
Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G.,
Leistner, E. and Floss, H.G.
TITLE Mutational analysis and reconstituted expression of the
biosynthetic genes involved in the formation of
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
biosynthesis in Amycolatopsis mediterranei S699
J. Biol. Chem. 276 (16), 12546-12555 (2001)
MEDLINE 21201076
FUBMED 11278540
REFERENCE
4 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
TITLE Direct Submission
Submitted (30-DEC-1997) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
JOURNAL
5 (bases 1 to 90445)
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,
Hutchinson, C.R. and Floss, H.G.
TITLE Direct Submission
Submitted (19-JAN-2001) Dept. of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
JOURNAL
REMARK Sequence update by submitter
COMMENT On or before Jan 22, 2001 this sequence version replaced
GI:2792340, GI:2832263.
FEATURES
source
1..90445
/organism="Amycolatopsis mediterranei"
/mol_type="genomic DNA"
/strain="S699"
/db_xref="taxon:33910"
423..1451
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01706.1"
/db_xref="GI:2792310"
/transl_table="MAPTGGGSGSRVYKTRQLEVALIGAGTLARLHLEAWLGAGAAV
RVYSDGRSRELAEEFGAKAGSLEALDGAADVICTPASHHEIATLTAAGCGVV
CEPLAASAEAEIIVTAERAGVRLYAHADVPAAPYARLHLEASGRIGEGALGRF
SFSAIRPRTVHASARSGGILLTDQLLHDLADLAHVFGDVVRVHACVQGGIATPAPEG
AVATGATVLTHASGALSQVYSRWTAATPPRVPAFVHVGSGGSVSDSEWPQVRVVD
GAGNFAVGGPSFVDEMRBFATAGGPEPRICAKDALAAVRITHAABESATGAV
ELPVRGAA"
CDS

```

```

1448..2215
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01707.1"
/db_xref="GI:2792311"
/transl_table="MKVAILSTPOAYAGALRGLPDVEVVAASWDAFEPVRAABAG
ARVLCYPPAAKETDLKAMIDAGDELTFASPACGEAEANVRKGIADGGIGELTIVL
GSVASDGVGLGAAPYLLDLDAVLGGPAAQVYAOITNIVLSGRIGESAAVITVYR
SQGVASFDCRRHGSATGLPVTFIGDQSVQDAGPGLGGERPELGGEDLEALMLKD
FLGADGPGPDQGAALTRFIIQAAYESAHTQGPDL"
2353..2619
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC01708.1"
/db_xref="GI:2792312"
/transl_table="MODRSVLGDPKQVEAAMSVTQVKIALSGGPAELAPAHRSVDSG
ALENTLKIRHGAGYEHFVHGGEFQAVDGCDAIAFRWSHRTKIAE"
2652..3842
/notes="ORF0; similar to Streptomyces Sp. Chop"
/codon_start=1
/transl_table=11
/product="cytochrome P450 monooxygenase"
/protein_id="AAC01709.1"
/db_xref="GI:2792313"
/transl_table="MTDAISFEVPWDRDKFPPAVFDSLREERFLAKMVYDPGHVGW
IVSSYELVREVLSDLRFHSCEVGHFPVTHQGVIPHTPLIPGMFIHMDPPETRRRK
LITGEFTVRRASRLIPRAEAAEQTEVNRKAGPADVMDVDFAKPLVLRMLGELGLP
YERDRYPAVTLHDAEADPAEAAAYEAGKVFDEVIERRRQPDQDLISLVTED
LTOEELRNIVTLFGAYFTTGALATGVFALLHHTDQLAARAEPEKLDAAEELLR
VYVNGQHYRTALDEDVKLEGLIKGDVTVYSLPAANDPDAKFGCPAELDIERDTSG
HVAFGEGHQCLQNLARIELRAGFTALLRAPPELRVLAVIDEVPRLKGSVPSVKKL
PVSWM"
3883..18090
/gene="rifa"
3883..18090
/gene="rifa"
/notes="rifa"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAC01710.1"
/db_xref="GI:2792314"
/transl_table="WRTDLIKLHVALLNATRFACKPAFADHRTVTYGDLEARTR
LAGHLAGLVGRHGRVAILCLGNKRVSTVESFALLRAGAVGPLNPGSATAELEHPTD
SGATVYVTDAAQARLELAPHVELLVTDGDDVEGASHYDELALESPAEADDLELDE
PAMWFTSGTTGRPKGVSTQRNCLMSVASCTVPPFGLSQDQDRVLPFLPFLSHLSHIA
CVLSATVVGASVRIADGSSADDVMDRLIEAESSTFLAGVPTVYHLLVRAARQRFSAFS
LRIGLAGGAVLGALRSFEETFGVPLIDAYGSTETCGAITWNPDPGAVEGSCGLAV
PGVDVRVDPDGLDVPAGEGEVWVGWMLGHNSEATAAAMRDQWFRIGDLAR
RODAGFTICGRKELIKRGGANHPGEVAVLRIVDGVADAAGVGPVPHDTLGEVPA
YVPGTGFDPAALEKCEQLEISAYKVPDRILEVAHIPRTASGKIRGLLTDEPAQUR
FRDLGFTSLAIVELNRNLTHTGLMPASAVDHTPAAALAAVRAELIGITQAAVEAP
VVAADPEPIATVGMACRLPGGVASPEDLRLVAVRDVAVSEFPFGDGLDLSLIDPD
RERACTSYVGGFGLHDAGEFDAGFPGLSPREAVAMDPOORLLLETSWLENAGVDP
IALKIDTGVFSLGMQGGSGGAPFELGFTVTGAVSASVSGSVVGLGLEGPATVV
DTACSSSLVAMHLAALQALGCSMALAGGVVTMATPFGSFVRQRLAFGRALCKAF
AAAADGTGWSGEGVVVLERLSRAGERHRILAVLRGSVNDQDGSNGLTAPNGLSQQ
RVIRALAAAGLAPSDVDVEAHGTTLIGDPIEAQALLATYGOERKQWLGLSKPSN
IGHQAAGAGVGVIKWQALRHETLPTLHVDKPTLEVDWSAGAILLFEARAWPENG
RPRAGVSFGVSGTNAHLILREAPAEVPAPELPVPLVVSARSTESLSQAEELA
SLLEGDVLTEVAGALVSRAVIDERAVVAGSREAVTGLRALNTAGSGTGGKVVWV
FFCQGTQWAGMGRLEILASPFVAFERIAECAALAPWIDMSLVLDLEADKAVLRQAIA
PACFAVGLAAAVWESVGRVDAVVGHSOGEIAAACVSGALSLEDAKVVALLRQAIA
AELSGRGGSAAVAGREDVVSRVGVVAANGSVSVIAGDAHALDATLETLSGEG
IRVRRVAVDYASHTRVEDIRTLAETLAGISAQAPAVPFYSTVTSWTRDAGVLGG
WYRNLRNVRFGAATALLGQHTVFEVSAHPVTVPGLSELDTGALTLRRDGLG
RLLASNGVFLVRGIDVMDWTWPAAGVMDLPTIYAFEHRYMLLEPASASGDPLLGT
VYSTPGSDRLTAQWRSRAQWMDGLVFNALVEAATRLGLDLAGTVPVGLVAVD
VVLPRGSEVOLIVGCEGQERRRPIEFVFSREADEPWRHAGHTLAPAAAAPVPAAP
GDATDVTVAGLRADADRYGHTHALLDAAVETVVGDDLLPSVMTGVSLLASGATVTFP

```



VAHALATTAAQLDHPAVVLASDRDLRADLAAFGPGVVTGTPTVDGKLAFLVFTGGSSOW  
AGMRELACTFFVFNRAEACHAYDAHLGRAPVREVVDDETRLDQMYTQGAFLAV  
ETALFRLFASVDPDLAAGHSIGELAAHVGSGVLDLAAAGELVAARGLMAQALPAGG  
AMVAQAAEDVAPLLDGTVCVAAVYNGSPDSVLSGAEDAVLAAGAARGLRKRLLP  
VSHAFFSLPMEGLDFRAVERLRYRAGTLFVSTLTGKLAALDTPGYVVRQVRDVA  
REGDAVTELAACGATYLELPGGALAAMALGTGGPQSCVATLRKNGAEPVGYLTA  
LAELHVRGTVDWMTALLDRPAGTVLPTAFYAFHERVWGTDETPSPASPVAAAPAE  
PVHDVPALVRESVAVVGLHRDAAADLDRSFKDLGFLDSLGAALKRNLDRDFTGVDLPS  
TLIFDYPNAUJADHRAELRERPAVAPVVRDVSDEPIAVGMSRTRPLPGADSPFE  
LNLVAEGDAISGFFVDRGMDLGLYHPDPHAGTSYTRTGGFLHDAQFADGALFI  
SPREALMDPQORLLIETSWEARERAGVDPDSARGSDVGVFTGI VHHIDYVTRLEVPE  
DYQGTMTGTASASVAGVAYVFGEPATVDTACSSLSVAIHAAQAALRQCEGSM  
LAGGATVMA SPDAFLFESRQSLSDGRCKVAEGADGTGMAEGVGVVVLRLSVARE  
RGH"

ORIGIN

Query Match 86.9%; Score 435.4; DB 1; Length 3900;  
Best Local Similarity 91.8%; Pred. No. 1.3e-37;  
Matches 460; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 1 TCGTCTTCGCTGCTCGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTCTCC 60  
DB 733 TCGTCTTCGCTGCTGCTGCACATGGCGGTGCAGGCCCTGCACCGCGCGAGTCCGCC 792  
QY 61 ATGGCGCTCGCGCGCGGTGATGTTGATGGGCACCGTTCGAGACGTTTCGTCGAGTTCTCG 120  
DB 793 ATGGCGCTCGCGCGCGGTGATGTTGATGGGCACCGTTCGAGACCTTCGTCGAGTTCTCC 852  
QY 121 CGGACGCGGGCTGGCCCCCGCGCGGTGCAGCGCGTTCGCGACGCGCGGACGCGC 180  
DB 853 CGGACGCGGGCTGGCCCCCGCGCGGTGCAGCGCGTTCGCGACGCGCGGACGCGC 912  
QY 181 ACCGCTGTGTCGAGGGCTCGGGCTCTCTGTTGAGCGGTGTCGAGGGCTCAGCGT 240  
DB 913 ACCGCTGTGTCGAGGGCTCGGGCTCTCTGTTGAGCGGTGTCGAGGGCTCAGCGT 972  
QY 241 CGCGCGCACAGTCTCTCGCGGTTCGCGGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 300  
DB 973 CGCGCGCACAGTCTCTCGCGGTTCGCGGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 1032  
QY 301 AACGGCTTGAACGCGCGCGCGCGGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGT 360  
DB 1033 AACGGCTTGAACGCGCGCGCGCGGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGT 1092  
QY 361 GCGCGCGACGTCTCCATCGAGCTCGAGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 420  
DB 1093 GCGCGCGACGTCTCCATCGAGCTCGAGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 1152  
QY 421 CTGGCGCACCGATCGAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 480  
DB 1153 CTGGCGCACCGATCGAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 1212  
QY 481 CCGCTGTGCTCGGGTCCGGTGGTG 501  
DB 1213 CCGCTGTGCTCGGGTCCGGTGGTG 1233

RESULT 5

A69718  
LOCUS A69718 5676 bp DNA linear PAT 07-MAY-1999  
DEFINITION Sequence 1 from Patent WO9807868.  
ACCESSION A69718  
VERSION A69718.1 GI:4774327  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 5676)  
AUTHORS Schupp,T., Toupet,C. and Engel,N.  
TITLE RIFAMYCIN BIOSYNTHESIS GENE CLUSTER  
JOURNAL Patent: WO 9807868-A 1 26-FEB-1998;  
CIBA GEIGY AG (CH)  
FEATURES Location/Qualifiers

source

1. 5676  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 66.8%; Score 334.6; DB 6; Length 5676;  
Best Local Similarity 80.3%; Pred. No. 4.8e-27;  
Matches 407; Conservative 0; Mismatches 94; Indels 6; Gaps 1;  
QY 1 TCGTCTTCGCTGCTCGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTCTCC 60  
DB 3897 TCGTCTTCGCTGCTGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGCGAGTCTCG 3956  
QY 61 ATGGCGCTCGCGCGCGGTGATGTTGATGGGCACCGTTCGAGACGTTTCGTCGAGTTCTCG 120  
DB 3957 ATGGCGCTCGCGCGGTGATGTTGATGGGCACCGTTCGTCGAGTTTCGTCGAGTTTCGCG 4016  
QY 121 CGGACGCGGGCTGGCCCCCGCGCGGTGCAGCGCGTTCGCGACGCGCGGACGCGC 180  
DB 4017 AAGCAGCGCGGTTCGCGCGGTGCAGCGCGTTCGCGACGCGCGGACGCGC 4076  
QY 181 ACCGCTGTGTCGAGGGCTCGGGCTCTCTGTTGAGCGGTGTCGAGGGCTCAGCGT 240  
DB 4077 ACCGCTGTGTCGAGGGCTCGGGCTCTCTGTTGAGCGGTGTCGAGGGCTCAGCGT 4136  
QY 241 CGCGCGCACAGTCTCTCGCGGTTCGCGGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 300  
DB 4137 CGCGCGCACAGTCTCTCGCGGTTCGCGGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 4196  
QY 301 AACGGCTTGAACGCGCGCGCGGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 360  
DB 4197 AACGGCTTGAACGCGCGCGCGGTTCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 4256  
QY 361 GCGCGCGACGTCTCCATCGAGCTCGAGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 420  
DB 4257 GCGCGCGACGTCTCCATCGAGCTCGAGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 4316  
QY 421 CTGGCGCACCGATCGAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 474  
DB 4317 CTGGCGCACCGATCGAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 4376  
QY 475 GAAACGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 501  
DB 4377 GAGACGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4403

RESULT 6

AF263912 123580 bp DNA linear BCT 24-MAY-2000  
LOCUS Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,  
DEFINITION complete sequence.  
ACCESSION AF263912  
VERSION AF263912.1 GI:8050835  
KEYWORDS  
SOURCE Streptomyces noursei  
ORGANISM Streptomyces noursei  
REFERENCE 1 (bases 1 to 123580)  
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strim,A.R.,  
Valla,S. and Zotchev,S.B.  
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in  
Streptomyces noursei ATCC 11455: analysis of the gene cluster and  
deduction of the biosynthetic pathway  
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)  
MEDLINE 20334850  
PUBMED 10873841  
REFERENCE 2 (bases 1 to 123580)  
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,  
Valla,S. and Zotchev,S.B.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim



```

FEATURES             source
N-7489, Norway      Location/Qualifiers
1..123580
/organism="Streptomyces noursei"
/mol_type="genomic DNA"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement(46..783)
/gene="nysF"
complement(46..783)
/gene="nysF"
domains on PKs"
/function="presumably post-translationally modifies ACP
domains on PKs"
/note="putative 4'-phosphopanthetheine transferase"
/codon_start=1
/transl_table=11
/product="NysF"
/protein_id="AAF71762.1"
/db_xref="GI:8050836"
/translation="MIELILPATVATAAAYDDRPDGRLLSSRREVIARAVESRORE
FTTVRLHRLALPGLHDPDRAILLNRRGAPQWPGIIVGSMTHCAGYRAAAVSPAELSA
AVSIDAPNGPLPAGVLNALPESRPHILVALAHRDDVNDWDLFLFSAKESVFKWYTP
LTQRELDSEAEIVDPTQGAFTARLLVPGPLGGRRVTVPPGRWHSTPALLTTAVHL
PAPTRDRHRLTHLYNSLPRTFG"
complement(867..2684)
/gene="nysG"
complement(867..2684)
/gene="nysG"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="MASPDLLLEERTAPPRARRLVGLLRPHRRSVALLAVSMGVGIVL
NAFGLLGRVTDILADVGVPAPGIDFAIGRLLVLLVLLVYVVASLFLMAQGR
LIVASVNTHELARDAREKILRLPHRFDPQAGELLSTRTDINDLQTLQTLAE
LITSPILLTMLVLSISLAVLMLLSVPSALIAARISKRAQPHYAQAQWANGTL
NAHVEECTGHALIGDRRAAEERFDACNDVIRAAKQAFAGMEVMMFVNL
GYVLVAVGAWKVIINGTLTGDOVQAFILYARQSPQIVETIASVAGRLQSGTSAQRPV
TLIDAPQAPDPLRPFGTPARAGRVETDVSFRYSPDPTLLENLSLTVPEGSTVAIYG
PTGAKTTLGNLMLRFPEDSGRILLDGTDTATWTRDRLSRFGVLQDWTWFGGTIA
EINAYAGPACRAIDIBRAARATCADRTLPQGVDTVLDSEGTVSAGEKOLLTVAR
AFLARPVLVLDSEATSSVDTEVLIQAMNSLRAGRTSFVIAHRLSTIRADLIIVM
DAGRIVQGTQDLQCLQGLYARLHAARKHTPTTGAAG"
complement(2662..4416)
/gene="nysH"
complement(2662..4416)
/gene="nysH"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translation="MLKLRLRAQLRPYAWATAALVALQVLGTLILLPTLGAALIDQ
GVNRGDGRITELGVMMGVVALVQIAALGAALAAATATATANGDRLSALFRLLDPS
AREIRFGTPLLTRSYNDVQVQNLAQTFGIVVCAPIMLCLGSLVLLARQDVLALL
LVALVVAVCGFLLIARMGTLVQRMQLTDLGRLLREALTIGRVVRSFVRDDBHERA
RPAQNTDAFLVSVRVLGLIATMLPVILLMNGFTVALLTGSRIIDAGRMPIGLSLA
LLSLILMSVMVLAFLVPSVRVACRIAEVLDTGSSVAPPAAPQVRGPAGRI
ELCRAGIYRGAEEPRKVDLITVEFGRIVAGTSGKTLTLLMLVRLADATAGV
RVGDTVRELTAATIAAAGVFPVQRFSGTVASNTFRPDATDEELMADQVQDA
ADFVAMPDGLDAITLQGGGVNGGQRFQSLRLSARLLERPEIYDFDQCFSLADQATDA
ALRTALVPYTAGVITVAQRIISAGRDADRIVWLDGRVVAQGTHTDVLRLTSPTYREI
ALSOLTEEAAGLAGRP"
4714..5748
/gene="nysDIII"
4714..5748
/gene="nysDIII"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIII"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITQDGSYLAEHLISQGYQVWGLIRGOANPRKSRV
SRLASELDFIDGLMDQSLVSAVDVQDPEVYNLGAISFVPMNSWQAEALVTEVNGMG
VLMLERAPMVSGLSRTSVSPRGQIRFYQASSEMFGKAEFTQRETTLTFHPRSPYG
AAKAYHITRNTRESFGMYAVSOMLENHSPRGQEFYTRKISLAVARLKQGLDKL
ALGNLDAVRDMWYAGDYVRAMHLMQDQDGDYVIGTGQMSHVSADAVRIAFEHVGLNW
EDTVVLDPLVRPAEVEVLCAQDLGKPKDVPDFTPLMRMWDSDLAQVSRNQ
YGDVLLAANW"
5930..34363
/gene="nysI"
5930..34363
/gene="nysI"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NysI"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translation="MDNEOKRLDYLLATADLRTRRRVHKLESAAQEPVAILIGMTCR
YPGVRSPEDLRMWVEAGEHGVTPPTDRGMDLEALAAATASGGLFHDAPDFDADF
GISPREAVMDPQORVVLESANWEAFERAGIDPITSVKSRCGVFGMAQDYRVGPADG
ABGQLTGNTGTSVLSGRISYTFGTGVPATVTDACSSLVAVHLATQALRAGECTIAL
AGGVTINSGPFTFEMRGQGLSADGRCSFGDTADGTGWABGGLILVLERLSDAVRN
GHEILAVRGRTAVNQDASNGLTAPNGFSQQQVIQQALVNARLADIDVVEAHEGTGT
TLGDVPEAOALLATYQONRPAIRPLILGSKVSNLSHTQAAAGVAGVIMVMAMEHGTL
PRTLHAEPTHVDMSQGAVALLLTDTTWPATGAPRAAASVSGISGTNAHTITIQAP
EPQEDAEATQDDAAGSTPATPVVPGVPSLLSGRTPDALRGQAALRALRDTRGRP
DLDLAHSLATTRAGFEHRAVILLATDHPALDPAULDGLTALADADDPAAAPAMITGTRAE
RLVATFGQAQRILAGRELAARFPATADALDAADFTPHLDRLEVLWMTQOALL
DR7AYAQPLFAVEVALYRLTESFGVRPDLHAGSVGEIYAAHLAGVLSLADAATLVA
ARGELMALPDGGAMIAVOASEADVAPLLAGEHDOVAIAAVNGSPAVVLSGAETVTA
LAEQALADGRKTRLYSHAFSPLEMEPLMDAFAVVEDLTLQPLPLPVSNLNGKPA
TVAQTSADVNDVHRVHAFADGIDMLARHDTTAFLELGPDGVLSMAQCLDAADA
DAVTLPALRAGRPEEHTLTALAGLHVHGTALDWTGCFAGTGARRTDLPTFAFORRY
WPKLQSGTADRLSEVGLAAHPILLSAVSLADAGTLLTGLSRQTHPLADHTVRG
WPLTGTAFLVLAAGDVECDRVEBLTLAAPLLPEQGVQVQLMTGNPDVSGRRIT
VNVHARPDGDTPTAHATGLTADAROLPASSEGQTPLAGDHPALDAQWPP
AGAEPLDGHYDLADGGEGYGVFOGLRAAWEVGVDVVAEVELPEAGRSDAAGFL
HPALLDALHAAPFTGLGERGGLPFSWEGVSLHAGGATTLRLLTPVADDAALITY
ADGTGAPVLSVDSLVRSVATQQLDTAAAVARDALFLDWTPTVQPTATDPGFVALLGA
DPFGLLTHAGFADAPAYPDLAALAAADGPVTTVLSLADTGGDDAADPARSAHCAAE
ALAAVQTLQDHHERRFAAARLVFTRGATVGRDVAATAVWGLVRSQAQENPGCFALVDL
DPDCAVGAALVAALVSGEPOLAVRGDVLRLVRLRPLTEWGAGADGTGGVGDGSG
VPSGEGAVLVGTGGLGVLARHLVAEYGVDRDILLVRSRGERAVGAGELVAELAGV
GARVVRVACDVDRAAVVELVGHAVSAVHAAGVLDGVMGALTGERLSAVLRKVD
AVMHLHEATRGDLDTAFVFSLAGVFGSPQANYAANAFDLDMTRRABRGPLGS
LAWFQWEGGMTGTLTDVDAERLARSGVPPLSVAAQGLADFAADVAGTATCCVPVRLD
QVALVIGHATGDDVADRAPRDLGDSLTAVELRNRLNTVTLGLRLPATLVFDYPTVRH
LATVLLDGLLDAEVATQPAVADDPVIVGMACRYPGVSSPDLNRVLTGEGT
DAVSGFPTNRGDVRESLHDPDPHFGTSYTRSGGFLHAGDFPGFPGMSREALATD
SQQLLLESSWEALIERAGIDPVLRSGTGVFAGVMSYSDYSAMLASPFEGVQSGSS
PSLAGRVAAYRTLLEGPAVTDTACSSLVAMHAWMALRSGECALAGGVTWMTPT
AVFVDGARQGLSPGRCKFADADGQVMSGEGVGLVLERQSDAVRNHGLTILAVRG
SAVNDQASNGLTAPNGFSQQQVIRQAASGLTAGDVVDVVEAGTGTGLDPTIEQAQ
LLATYGRDREPERILLGSKVSNLSHTQAAAGVAGVIMVMAMEHGVTPTLTHVDAPS
SHVMSRGAELLSEQAANWPTGRRVRAGVSGISGTNVIVIVEQAGAKALAAAGA
ARRTPGVLLSGRSALRGQAARLHGLHQAOPDAELVDVALRFRFRFEBQRAA
VVAQDQLTLASLALADPDPVAVEGEAARGRTAVTLFTQSSQRAAMRGREFORAA
PEFAADPAVCAVFDPLRLREVFEAEDSDEAALIDETGCTQALPFTVLEAFVFLQ
VESNGVRDFFVAGHSIGRIAAHVAGVLTLEDACRLVAARATLMQALPTGAMTIAOIA
TEDEIAHLDLDTVALAANGVQSVVSGDEEAETIATFAERKTKRLRVSHAFHS
PRMDGLDAFRIVAEGLTYRAPRPLVSDLTGRRADDDEVCTAEYVWRHREAVRFAD
CVTRLRAGATTFLGSDGLLTAAEDTLGDDHDAEPLMRLRAGRASELAATAALAR
LQVRGVDVDMAAVYLAGTGARTDLPTVAFQHAYTWPQLPTPAAALAAADPADQQLWAA

```

VERGDARELADITLIGAEODLTPLDLSLLPALTSWRRGNQEKHLDDTLRYRVEVTRLSKP  
TAPVLGDGTWLLVASDADADOPALDGLADALSGHARVRRLDLDSDCADRAVLAERL  
ARTADVDRATQVLSVLPDIDERDADDCPELTRGLALITVALQALADTQAGRLWTATRG  
AVSTNPADVPVHPVQAAGWGLGRGVALHEPRLWGLVLDLPQVDFERAGORGLAIGLVK  
DAPDGEDVALATGVRGLRVRHTVEALPTAAEFATGTATGLITCGTGLGAEVAVRL  
ARAGAHVLVTSRRGPDAPGAELAELEGVPSVWACDVADRDLAALVLTALPEE  
LPLTGVTAGVHPDITLTAEPAFLTAALAGAAHLDALARELDLDFVLPFESI  
AGVMSGNSQASANAVLDALHRRARGLAATSVAMPWAEAGMAADAVSETLRR  
QGLGLDPAWTELRRAVVRQDVTVVADVDMQRYAPLFTSARPSALAGIPEVRL  
AADETEQDAGASEVTVRVALAPAEQRLDLVRIESATSVLGHSSADAVPEGRAP  
RVGDFSLTAVLRKRLGATGLSLPSTWFDYPTPLLELAQYLRAELIAGVLEAGVP  
ATGQADDEPIAIIAGACRPGVSSPEQLMDLVASGTDAISEFPVNRQMOTGLHDPD  
PRPGTITQTCGFHEADPFTFPGISPREALVMDPOQLLETTFESFERAGIRP  
ETRLTSLTCTFVSSVQYGLGAGDTEGHMTGSSPVLGRSLVTFGLEGPVATVD  
TACSSSLVHALIACQSLRNGESNLAVAGATIMTPNPFIAPSORALAKDRCRKAFS  
DDAGMTLAGVGVILVERLSDAQRNGHPVLAVRGSAINQDQASNGLTAPNGPQQOR  
VIRQALANARAPGDIDALEHGTPLGDPTEAQAALFATYGRDRDPESALLIGLVKS  
NIGHTQSAAGIASVIMNMALRHSLEPPTLHADAPSHVDWAGTVRLITQARAWPET  
GRPRAAVSFGISGTNAHVLEQAPVADTPAERPAVAPVPIAGVWPVWTVARSAA  
ALRGOAERLLHAETVTGALPAAGPLDIGLSVSARARFEHRAVVPVPAAGTDLAALR  
AVATDGPSPVARGVADVEGRVFPVPGQSGWQWGMGSQLDESAPFAPRIAECHAL  
AFTDMSLDVLRGVVGAESLERVVVPQSPAFVMSLAALMRSGVLPDPAVVHSGQ  
ETAAVVSGLSLRGRVALRSQAIGRALAGRMMSVALSDVLEPRLEFEGERV  
SVAANVSPRSVAVAGEPEALDHALHARLTADDIARRIADVYASHQVEDLHEELLEV  
LAELAPRTSVFVEFTVTDGMDLTARMAGYFNRNLGRVRFADAVADLILAEYRAFV  
EVSHPVLQSWAEFEATDEAGVPAVAGTLRRDQGTDRLLSAAEFVVRGVVDVWAGL  
FRGTGASRLDLPYAFQHEHLNVPAPPAVAAADPDDAAFWTAVEDGDVSLTALTAGL  
TIEDSVAAVLPALTSWRARRDRSDVDAWRVYRWKPLGGLTLPHPSLTGLTLLVYADG  
IDDTVAGALETYGAERVLRLDDECDVRLRERLAGAEDVTGIVSVLAAERATDAV  
PGTSLVLGTALTVALTQALGDABIDAPVWALTEGAVSTGRADELTPAQVQTGIGMT

Query Match 66.3%; Score 332; DB 1; Length 123580;  
Best Local Similarity 79.0%; Pred. No. 3.2e-27;  
Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1 TCCTCTTCGTCGTCGTCGTCACATGCGCGGTGCAGCCCTCGACGCGCGCGAGTGTCTC 60  
Db 35023 TCCTCTCTGCTGCTCGCCCTGCACCTGGCGTGCAGCGCTGCAGCAACGCGAGTGC 35082

QY 61 ATGGCGTCGCGCGCGCGGTGATGGGCACTGGGCACTGGGCACTGGGCACTGGGCACT 120  
Db 35083 ATGGCGTCGCGCGCGCGGTGATGGGCACTGGGCACTGGGCACTGGGCACTGGG 35142

QY 121 CGGACG 180  
Db 35143 CGGACG 35202

QY 181 ACCGGCTGGTCCGAGGCGGTGCGGCTGCTCTCTGGTGGAGCGGCTGTCGAGGCTCAG 240  
Db 35203 ACCGGCTGGTCCGAGGCGCGCGCGGTGCTGCTGTGTAACGGCTCTCCGACGCC 35262

QY 241 CGGGGACACAGGTCCTCGCGGTGGTTCGGGGTTCGGGGTTCGACTCGAGGCGCGTGC 300  
Db 35263 AACGGCCACCGGTCGTCGCGGTGGTTCGGGGTTCGGGGTTCGACTCGAGGCGCG 35322

QY 301 AACGGCTTGAACG 360  
Db 35323 AACGGCTTGAACG 35382

QY 361 GCCCGCGGACTGTGCACATCGGACGTGCAGCGGTGGAGGCGGCAACGCGACGCGAC 420  
Db 35383 AACCGCGGCTGGCG 35442

QY 421 CTGGGCGACCCGATCGAGGCGCGCGGTGCTGGCCACCTACGCGCAGAACGCGGAAACG 480  
Db 35443 CTGGGCGACCCGATCGAGGCGCGCGGTGCTGGCCACCTACGCGCAGAACGCGGAA 35502

QY 481 CCGCTGTGCTCGGTCGGT 500  
Db 35503 CCGGTGTGCTCGGTCGGT 35522

RESULT 7

AX211739  
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 35 from Patent WO0159126.  
ACCESSION AX211739  
VERSION AX211739.1 GI:15523950  
KEYWORDS Streptomyces noursei  
SOURCE Streptomyces noursei  
ORGANISM Streptomyces noursei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
1 Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, F.,  
Stroem, A.R., Walla, S., Ellingsen, T.E., Sletta, H.V. and  
Gulliksen, O.M.  
Gene cluster encoding a nystatin polyketide synthase and its  
manipulation and utility  
Patent: WO 0159126-A 35 16-AUG-2001;  
Norges Teknisk Naturvitenskapelige Universitet (NO); STIPELSEN  
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);  
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich  
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);  
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Walla, Svein  
(NO)  
Location/Qualifiers  
1. .125401  
/organism="Streptomyces noursei"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1971"  
/note="ATCC 11455"

Query Match 66.3%; Score 332; DB 6; Length 125401;  
Best Local Similarity 79.0%; Pred. No. 3.2e-27;  
Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1 TCCTCTTCGTCGTCGTCGTCACATGCGCGGTGCAGCCCTCGACGCGCGCGAGTGTCTC 60  
Db 35431 TCCTCTCTGCTGCTCGCCCTGCACCTGGCGTGCAGCGCTGCAGCAACGCGAGTGC 35490

QY 61 ATGGCGTCGCGCGCGCGGTGATGGGCACTGGGCACTGGGCACTGGGCACTGGGCACT 120  
Db 35491 ATGGCGTCGCGCGCGCGGTGATGGGCACTGGGCACTGGGCACTGGGCACTGGG 35550

QY 121 CGGACG 180  
Db 35551 CGGACG 35610

QY 181 ACCGGCTGGTCCGAGGCGGTGCGGCTGCTCTCTGGTGGAGCGGCTGTCGAGGCTCAG 240  
Db 35611 ACCGGCTGGTCCGAGGCGCGCGCGGTGCTCTCTGGTGGAGCGGCTGTCGAGGCTC 35670

QY 241 CGGGGACACAGGTCCTCGCGGTGGTTCGGGGTTCGGGGTTCGACTCGAGGCGCGTGC 300  
Db 35671 AACGGCCACCGGTCGTCGCGGTGGTTCGGGGTTCGGGGTTCGACTCGAGGCGCG 35730

QY 301 AACGGCTTGAACG 360  
Db 35731 AACGGCTTGAACG 35790

QY 361 GCCCGCGGACTGTGCACATCGGACGTGCAGCGGTGGAGGCGGCAACGCGACGCGAC 420  
Db 35791 AACCGCGGCTGGCG 35850

QY 421 CTGGGCGACCCGATCGAGGCGCGCGGTGCTGGCCACCTACGCGCAGAACGCGGAAACG 480  
Db 35851 CTGGGCGACCCGATCGAGGCGCGCGGTGCTGGCCACCTACGCGCAGAACGCGGAA 35910

QY 481 CCGCTGTGCTCGGTCGGT 500  
Db 35911 CCGGTGTGCTCGGTCGGT 35930

RESULT 8



SNA132222 Streptomyces natalensis pms1 gene. 20394 bp DNA linear BCT 24-APR-1999  
 DEFINITION Streptomyces natalensis pms1 gene.  
 ACCESSION A0132222  
 VERSION A0132222.1 GI:4678702  
 KEYWORDS pimaricin biosynthesis; pms1 gene; polyketide synthase.  
 SOURCE Streptomyces natalensis  
 ORGANISM Streptomyces natalensis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 Aparicio, J.F., Colina, A.J., Ceballos, E. and Martin, J.F.  
 AUTHORS The biosynthetic gene cluster for the 26-membered ring polyene  
 TITLE macrolide pimaricin. A new polyketide synthase organization encoded  
 by two subclusters separated by functionalization genes  
 J. Biol. Chem. 274 (15), 10133-10139 (1999)  
 99214571  
 10187796  
 2 (bases 1 to 20394)  
 REFERENCE Aparicio, J.F.  
 AUTHORS Direct Submission  
 TITLE Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.  
 JOURNAL Real, 1-Leon, 24006, Spain  
 FEATURES  
 Location/Qualifiers  
 1..20394  
 /organism="Streptomyces natalensis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:68242"  
 1..20394  
 /gene="pms1"  
 /gene="pms1"  
 /function="pimaricin biosynthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /product="polyketide synthase"  
 /protein\_id="CAB41041.1"  
 /db\_xref="GI:4678703"  
 /db\_xref="GOA:Q9X993"  
 /db\_xref="SPTREMBL:Q9X993"  
 /translation="MSNEEKREYIKRAIDHETRQQLDTEAKQREPLAIVSWACR  
 FPGFVPEELWEILLRVDVAVSFFPRNRWDLDLHSDPAHQGTSTAREGGHILDA  
 HEDPFGFISPRFLAMDPOORLLLETAMEAVEAGIDPESLAGSTGYREGVCHGG  
 YDAERGRADRVGGHLLTGNHISASGRISYVLGEPALTVDTACSSILVGLHMH  
 ALRDECMALVCGATVMSFTQMFVFSRQRLAPDGRCKPFAAADQGTGWSGVGL  
 LVERLSADRVNGYFVLAVLKSANVDGASNLTPNGPSQQRVIRQALTCGLAASD  
 IDAVEAGTGTLDGPEFAHALLATYGOORADRPCGLGWSKNIGHTQAAAGIAGVM  
 KMLAMRHGHLPRTILDHDEPTGHVDWSEGNRLILAEPPWPSPAPRRAAVSSFGISG  
 TNAHVILEQAPAEAPAEAPAPRPGALPMTLSARTKAEAGRAQRLGRHLRDRADLE  
 PAAYHALADRTILMEHRAVAVAGDRBEFLRGLDALAAGRTANGLVGVAVKAAAFLL  
 FAGQSORPGMRELLHAHPVFATADAVCAELDPHLDRPLRDIVPAEBSGAALIL  
 QTAQTOALFALETALFVLSMGVAPFVAGHSIGELTAHVSGVLTLDHDAARLVAA  
 RGTLMQALPAGMVAQATDETRERLAGHEDHVALAANGPDSVTISGDEQAVTEI  
 AAHWEAQRRYTRLRVSHAFHSRPHMDMLEDFRVRAGLTTHAPRIPVVSTVGTALAT  
 EDELRSPDYVYRQVRVTRFCAAVITLAEAGVTTFVEITGGVLTTPMWQCLITLLEP  
 VLPLLTRGRTETVALTGEGVATFVHGVVDPSAPFAPAGTSRADLPTAFYARQWYLL  
 DPADHDEGAEEAAGEAGFWAAVEREDQLBSAVLADGSEADSLGSLFTLSSWR  
 ORRTCAADRFSYTHWPRATASGPTATGHWIVLPEGGTDDPWTARLLDALNDQGL  
 HTDVRLELPADHPEPWANGHVPDGVCLIALDERKTRVSCPPYRRAGLAATTNAAREPGA  
 QIQAPLVCMTVGAVAHDEHAKGLPQATQWGLGRVSALESPPSGWGLIDLPDLNDR  
 ARRAGLPAEGRRPERSPLRDRIGALPPAEQKAFITWREAAARLVHPSPDPTDAQR  
 AFRRQGGDSMAVLRNLSAATGLRIPATLLFPHPTPLAAACLSLVLGAAGPATV  
 VQASTALDEPVAIIGMACRPGGVHSPALWLLAEGGDAITPMRADRGWDLRLVH  
 PPDHQCTSYARGGGLDGAADFDGEGISPREALAMDPOORLLLETWETVEQAGID  
 PESLRSGSTGVFAGINTODYGTALDAQDEAGGRTLTGNMVSVSGRSTVYTFEGEPA  
 LTVDAACSSILVALHMAAQLRQECSLAVAGGVVMTAFSSFPFARQNSGLAPDGR  
 KPFAAEDSGTSGEGVGLLILVERLSADRVNGHVLAVRGSANVDGASNLGSLAPSGP  
 SQQRVIRQALANARVAASEVDVAEHGTGTTGLDPIEAQALLATYQGERPLLLGVKS

NLGHTOAAAGVAGVWKMVLMRHMGLPRTLLHVDERTGHVDWTAGAVELLTHTDWPET  
 GHPRRAVSAFGISCTNAHVLELEPAEQPLVEQPSAEADPAAPVTPRTPTASDGT  
 PLLLSAKSESALRAQAARLHSHLERDPAIRLITDAITMTHTTAFARAAVRAADHEA  
 ALRALITALAGEADPAVDGTGHTGRDVAFFSGQSQRIKMGRELSGYSVPFAEAFDT  
 VCAALDELDRPDRVVRGDEDELLNRTVYAQGLFAIEVALFRLESVGPVPHYVAT  
 HSGVEITAAHVAVFSLADALCAVAAARGLMAGPAGMAAIIATEDVTEPHLADSV  
 SIAAVNGPSSVVSAGAEHVLSTIAAFHFGAGKTRLRVSHAFSLMDPLMDPRAV  
 AEGLTGPELAVSVNTVQOLATPDQLTPTVWTHVAAVFAADIGALCAEGVTRF  
 LEGLPDGVLALARESAPDDAVCTPVLKDRSEATLLAALTHLHVHGTEDIDWTFALA  
 GRDAHVDLPTAFQHQRFWPTPDHRTDGLGAVGLEATGHEPLSLMELDPDGLLIF  
 TTRLSLQTHPLVAGHVMGSLVLPCTFAELARAAADVGCDDVDELSTALAPLVLPEH  
 GGVLQLRVPADASGRRTLTARSAGEDGRPMVQVHATGVLAGESTPEFCYDPHTE  
 SWPADAPAVELSLGLYDPFAAHGPDYGHFQGLTAMRRGDEVFAVPAALPAEAEES  
 AYLHPALLDAAHVAFNGVDGVVPSWESVALHATGASAVRTRVVRHSGDTSVD  
 VADTGTGPASITGTLVRAVSADQLAGGADPAVDALFRVQMPVLELPAGAAVAT  
 LGSLAGAPDGTGFDLASLRSGRVAGAVLVPVAGAGVAVDDVVVGHATHATRALDL  
 ARSLWADRFASRLVFTVRGAVSADLAGAAVMSLVERSLRSGFLGLDLDODAE  
 LALVRLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGCTGLGRMA  
 RLHVVEGRNLLVSRGPAABGAELVTLRHSGEAVAEACDVTDAAAADVLVAR  
 HRI SAVHTAGVLDGCVESLTPRLSAVLRPKVDAANLHEATDLDLDAFVVPSSV  
 AGTISPGQANYAAGNAFLDALAHERRAAGLPAASLANGPWSRDCMGTGLTDVSSA  
 SPGRHARTPRITGRGLFDALAAAGDAHLIPVRFMDASRAQGEVPEPLRGLLITRARR  
 SAVGSAAGLGLVRLSGRTVRRERVDLIVRAQIVAVLGHANPETESTRVPDGL  
 FDSLTAVELNRNLNATGLRSLTAVFDYPTDLRVLSCGVDVSPPTDRGWDVESLYSPD  
 PAGAADDPVITVGMSCRYPGGVGSPEDLWRLVSCGVDVSPPTDRGWDVESLYSPD  
 EALGTSYTRSGFLHEAAEPDPFGMSPREALATDAOQRLLELTWAIERTGIDPA  
 SLRSGRTGVAGVMYDGLLYGDOFEGYRSGNSAASIASGRVSYTFGFGPAVTD  
 TACSSSLVALHAAQSLRSGCSLAVAGVVTMSPTPTTFVFSRQRLGSLDCRCAKPA  
 DAADGWMGEGVGLVLERLSDARRNGHRLVAVRGSANQDGSGLTAPNCPAQQR  
 VIRQALASAGLAVDVAEHGTGTTGLDPTIEAQLATYQERPEPRLILGSKVS  
 NIGHQAASGAAVYKMLVLMRHGVLPTLHVPSPSHVDSAGAVELTSEAEWPOQ  
 EGPRRGSVSGVSTNAHVILEOPGDAADADATVTDGALAVLSARNEALRC  
 QARLLSVAGSDALCARDICHSVLTGESSFAHRAVWGDORDALVRLSALAGEAD  
 AGLAEGSACGRTAFLFSQGSQRLGMGWELIYRVPFADAFDAVCAALDEHLERPLR  
 DVWGEDAELNLQAYAQALFAIEVALYRLAESWGMRPDFVAGHSIGEVAASVSVIS  
 FSLPDACALVAAQRLQQLSPSGAMMAIATEDVLEPLHAEFRAVVTRELSPGTPTPV  
 CAEDAVLAIAAHFAEGEKTRLRVSHAFSPGLMEPLAEFRAVVTRELSPGTPTPV  
 SNTGLRLAEPEQLAHADYVVRHREAVRFDGLOALRAEGVTRFLSGDGLVLSAMAR  
 ESASDVAVLPRBDRPEETALLGALQVLRGHVDMVTPVFPAGSGARWADLTPYAF  
 QHERFWSGVARPGDVRSGAGLHPGLGAVALAGSGLLFTGRLUSVSHPWALD  
 HVVLGSLVGTALVELVLRADAGCDLLEELTAAPLVLPAGAAVQVQVAGEPD  
 DAGRYPVSHAREGEPMTLHSGAVTSGAEVFPFDATWPPKGAEPVVDVADCDVLA  
 DEGLTGPAPGLQAAWKLGDVAAKLPSTGDDAYGLHPALFDAAHSAAGLGAEE  
 FVFWAGVSLHATGASHLRVIRIAGAGLVAIADTSGAPVASVSVIRPLSAR  
 QVQAADRALPKADVPVPLTDERVECTGPEGLPTYAOLDLSLEGAASVCTVLVAP  
 PSGAAGTVESHAATVWALEWQAWLADDFATSLVFTVTRGAAGFADLAAAVRGLV  
 RSAQENPRGLVMDGDADTTPQAQALATDEPELLVRGGEVLAARLVRAQSSHTVT  
 WDPSTVLTITGCTGLGRSVARHIVSEHVRSLILVSRGPAAGAGELVLAELRSGGA  
 EVLTEACDVTDAVADLVARHRTSAVHTAGVLDGCVESLTPERLAVLRPKVDA  
 WNLHEATRGDLDAFVVPSSVAGTFGSAGQANYAAGNAFLDALAYHRRAGLPAVSLA  
 WGPSQGMGTPLSDADVORIARQGMPLTVESGLALFDALCSAEPWALPVRDLA  
 ALRAQGPMTPLRGLIETRRSAAAASGIAQRLAGLSTABERREALDVVRQIATV  
 LGHAGPETIAPDRAFQDLGLDLSLAIELRNLLGKATGLRLPATVTFDPTVDAALAH  
 LDEFPAGTGTATETPLVPGLSLADDPVIVGMSRCPGGVSPASPELMRLVADGV  
 AVSAPFDRGWEDTIDTPREGAIATRSGLFHDAAEFPFGMSPREALITDAQQ  
 RLILLETWEALRAGMDPATLRSGRTGVFAGVMDYSTLLSREFEYQSSGAGSV  
 ASHVSVTREGEPAVTDTCASSILVALHAAQSLRSGCSLALAGVVTMSLPTF  
 VEFROGLSADGCKKAPADAGVGMWAGAGLILVERLSADRVNGHRLITVRGSV  
 NQDSAGNLTAPNGPQORVIRQALASAGLAAADVDAVEHGTGTTGLDPIEAQALLA  
 TYQGEREDPRLILGSKVSINIGHQAASGAAVAGVYKMLVLMRHGVLPTLHVDSPSHV  
 HATRSPELMEPLADPRAVADGMTYAAAPRIPVVSVNTGRPAETAEELCAEYVGHREA  
 MPHRSPELMEPLADPRAVADGMTYAAAPRIPVVSVNTGRPAETAEELCAEYVGHREA  
 APDAEASDSTGTPVALLSGRSASALRAQAARLLSRLDGDPGRITVDVAYSLATGR  
 SAPHRAVILANRDLHLSLSALAEHTEAPVAVQDRASGLKATFLPSQGSQRLG  
 MGRILYRYPFAEALDACAALDHLDRPLRDVINGEDAEELLNRTGATGGLFAIEV  
 ALPFLRSGVGRPDHLGHSIGETAAAHVAGVLSLPDCAVLAARGMLQQLPGSGAM  
 MATRATEDPPLHAEVSLAANGVSSVVSAGAEVLAALAHFEERKTRLRVS  
 HATRSPELMEPLADPRAVADGMTYAAAPRIPVVSVNTGRPAETAEELCAEYVGHREA  
 VRFADGVALREQVTTFLHGPDSLSALAEASDADSVLAVLVRNPEAPALLTA  
 LARLHAQGVTPDMSAAPAGTGRWVDLPTFQHERFWPSSGAAARAGVRSAGLSGAS  
 HPLILGAOVELLAGSGRLITGRVLSLHPTLADHVVLGSLVLPFGTALMELVRLADEVD  
 APAADEITFOTAIWPPQASHLDAAGVCFERPADGAYFPVFGRLAAWKLGTIDYAE  
 VALPEGTDGNAGVGLHPALFDAAHAAALLGCGEGTDEAAVFPFSWNGVTLHATGASRVVR







```
MOALPEGAMVSLSEATEBEVLLLEDRPDLRLIAAVNGPAAVIAAGEADDVTAVAEHF
DALGRRTRKLRVSHAFSLPMEPLAEPTAVSLRTLPQAPVPIVSVGLTGLTAVTEOL
TSPOYWLHDAHVRFDALRLARRRHLPRMGDPGLSAMQSCCRGHGRRRR
LRHARRVPRGPREAETLTAALAGLHTRGVAVQWEPYFGTGARADLPTIYAFQRRYY
PKSLPATGGDVRAAGLGNHPLITAANSVANSGLLTLGRSLRTHFWLADHAVRG
VLLPETALELAVRAGDAGGRVEELTLAAPLVLPFEGGVQVQVWVGSPPDAGRRV
SVHRPDGPPELPTQHAAGTLVAGEHHAGFDATVPFPADATPLDLOGFYARMAATGF
GYGLFQGLRAARMSGDVVAEVALPDSIDGAATESAAFGHLHPALLIDLAALHAAAFVD
LGEDARGLPLPSWDVTLHAGSASTVRVKLTLPDGDDAVALAVADTTGAPVASIGSLVL

Query Match      65.5%; Score 328.2; DB 1; Length 84985;
Best Local Similarity 79.8%; Pred. No. 9.1e-27;
Matches 387; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 1 TCGTCTTCGCTGGTGGCTGCACATGCGCGTGCAGGCGCTGCAGCGCGGCGAGTGCTCC 60
Db 75953 TCCTCTCGTGGTGGCTGCACATGCGCGCGTGCAGGCGCTGCAGCGCGGCGAGTGCTCC 75894

Qy 61 ATGGCGCTCGCGGCGGCGTGCATGATGGCAAGTGCAGAGCGTTCGTCGAGTTCCTCG 120
Db 75893 CTGGCGGTGGCGGCGGCTGCAGCGTGCATGATGCCACCCCGTCTCTCTGTCGAGTTCGCC 75834

Qy 121 CGGACGCGCGGCTGGCCCGCGAGCGCGCTGCAGGCGTTCGCGAGCGCGCGCGACGCG 180
Db 75833 CGGACGCGCGGCTGGCCCGCGAGCGCGCTGCAGGCGTTCGCGAGCGCGCGCGACGCG 75774

Qy 181 ACCGCGTGGTCCGAGGCGGCTGGCGTCTCTGTTGGAGCGGCTGTCGAGCGCTCAGCGT 240
Db 75773 ACCGCGTGGAGGCGGCTGGCGTCTCTGTTGGAGCGGCTGTCGAGCGCTCAGCGCGCGCGA 75714

Qy 241 CGCGGCGACCAAGTCTCTCGCGTGGTGGCGGCTGGCGGCTCAATCTCGAGCGGCGGCTCG 300
Db 75713 AACGCGCACCAAGTCTCTCGCGTGGTGGCGGCTGGCGGCTCAATCTCGAGCGGCGGCTCG 75654

Qy 301 AACGCGTTCAGCGCGCGCGAGCGCGCTGCCAGAGCGCGTGCATCGCAAGCGCATGCGCC 360
Db 75653 AACGCGTTCAGCGCGCGCGAGCGCGCTGCCAGAGCGCGTGCATCGCAAGCGCGCATGCGCC 75594

Qy 361 GCCCGCGGCTGTCCACATCGGAGCGTGCAGCGGCTGGAGCGCGCGCGCGCGCGCGCGCG 420
Db 75593 AACCGCGGCTGGCGCGCTGCCAGAGCGCGTGCATCGCGCGTGGAGCGCGCGCGCGCGCG 75534

Qy 421 CTGGCGACCGCGATCGAGCGCGCGCGTGTGCGCCACCTACGCGCGCGAGAACCGGAGAACG 480
Db 75533 CTCGCGTACCGGATCGAGCGCGCGCGTGTGCGCCACCTACGCGCGCGAGCGCGCGCGTGT 75474

Qy 481 CCGCT 485
Db 75473 CTGCT 75469

RESULT 11
AB070940 104326 bp DNA linear BCT 22-JAN-2002
LOCUS
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION AB070940
VERSION AB070940.1 GI:15823967
KEYWORDS
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1
REFERENCE
AUTHORS Omura, S., Ikeda, H., Iehikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
FUBMED 11572948
REFERENCE 2 (bases 1 to 104326)
```

Ikeda, H.

Direct Submission

Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of  
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1  
Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail: ikedamc.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242,  
Fax: +81-3-3444-6197)

FEATURES Location/Qualifiers

source

1. 104326

/organism="Streptomyces avermitilis"

/mol\_type="genomic DNA"

/db\_xref="taxon:33903"

complement(1..1134)

/gene="mcmA"

complement(1..1134)

/gene="mcmA"

/note="methylmalonyl-CoA mutase"

/codon\_start=1

/transl\_table=1

/protein\_id="BAB69185.1"

/db\_xref="GI:15823968"

/translation="MTGISTAADANARPRQLVGSQTRAVPVVFDLPTFAGLSDAPAA  
RGLVGRGVAVDSIDDMRVLFGLRLDRVPSLAVDAPAPAEPLLYQLVAEHGAPAQ  
RLAGAVRDVKEFLARPGVFPPLRPSLRILAMNYFAYCAELRWQCLSVCGHLAGT  
GADPALEVFTVAGGIEYLRAAVTAGLDVTVAPRLTCLVGGTGTGRENKGLRTRR  
VMTRTNERFGSGRTAPRLYVTRPDGLATLACTEAGATALVDRLERLGGTLTVL  
EHGLGOLLGPHGCDPATPRPHEPHQDSLDLQARQTERLAKLANRWVQPAVDALLR  
LRXAAQDDNVLYPMREALAAATVGEVCDLURSTWTGTSAT"

complement(1131..1886)

/gene="pptA"

complement(1131..1886)

/gene="pptA"

/codon\_start=1

/transl\_table=1

/product="phosphopantetheinyl transferase"

/protein\_id="BAB69186.1"

/db\_xref="GI:15823969"

/translation="MRAHEVADVPSVSLHPREALVQDAVESRRREFATGRHCARRA  
LASLGVPPGILLATPGCAPKPGAGCGCAITHCAGYRAAANAVARAGIARVGDADPRP  
LRLGDAVALPAEPQFQVKEHTARHPGVVWVKLLFSAKESYIKAWFPATGPLGFEEA  
LIDLPVPSVFRVRLPTALRAAEPELRDLHGLVVRHGIIVTAVVVDGPRDTPHAP  
SPYDRKHSAPTGRSADRPGRPTDRATDPTA"

complement(1910..2875)

/gene="simX4 homolog"

complement(1910..2875)

/gene="simX4 homolog"

/codon\_start=1

/transl\_table=1

/product="SimX4 homolog"

/protein\_id="BAB69187.1"

/db\_xref="GI:15823970"

/translation="MSQDRHDGSTAALGSPSHGRLIAVSDIHRHQRNDRIVDLHP  
GCADFNLIVAGDVAESVADITWALSRLDRFAKVIWAPGNHFWLTPCPDDELGERK  
YRHLVDICRALGVPTDPDPYVWVGAGPVAVPLFLLYDYFPLPGWASKHALRLA  
DDAGICTDRELLHPDPYPSREAWCARVRATRLDALPDLPTVLVNHPPVLRPT  
RVLHRPEFLWCCTTATADWLRYRAAVVGHGLHIPTIECDQDPHEVSLGYPREW  
ORYGRPDWFOQLITAPAPAPVPVFPAPPGRHTASADRPP"

3213..3971

/gene="olmC"

3213..3971

/gene="olmC"

/codon\_start=1

/transl\_table=1

/product="thioesterase"

/protein\_id="BAB69188.1"

/db\_xref="GI:15823971"

/translation="MSANRVGDLNLYWRQHPAPTATKLYCLPHAGGSASFVLPVSKA  
LAPGVDLAIQYQRODRREKCTDNIAELADAVTAVLELADRLPLVLFHSGWATLA  
FEVALRLERKIVPLALFASGRAPSRHRDSVHLRGDDBITLALKTGSTRSEVIGD  
EELIRWVLPAIRSDYKAAETRYAGGARLTAP IAHAFGVEDPRVSLDEAQAWADHTTG  
HFELRSHPGGHFYINDQAPRVEIARVSSVAS"

4643..7438

/gene="olmR1"







PRRTSLAHGSGLRDLRLPLPODDRTTTLLENLRAEVSALPSADAVPTTKAFK SLGFSLSMAVDLNRSLTILTVGLRSATLIVFDPNTRVLAHLLTKRDLDTAGATGAL LALRGLEASPMNDPADRGAMATRLRLVLLATLEAGDEAAAGDSLDOSVASELSV LDDDFGLS"	Db	32371	CGGAGGGGGGGCTGCGCCCGGAGCGCGGTCAAGAGCGTTCCGCCAGCGGCGACGCG 32312
complement (11959, .26481) /gene="olmA2"	gene		
complement (11959, .26481) /gene="olmA2"	CDS		
/note="SAV2898 oligomycin gene cluster"			
/codon_start=1 /transl_table=11 /product="modular polyketide synthase"			
/db_xref="GI:29606548"			
/translation="MDDDKLVLYLKRVTADLRKTRQVQLRAGRAEPIAVVSMGCRF PGIASPDLWCFTVRLTDTISGFPTDRGMRTHLGRDFRLAGGLADAGDAGLFG ISPRALAMPDPOORLITLTSWEVLRAGLDPTVRGADGGVFGVADQYGRDGEELL DOVKGLVLTSSVASGRIAYSLGQGPATITDACSLSLVALHLAVRALRSGECSF ALVGNVAWEQLFAMAEQGMAGDGRCKAFAMAADGTGMEGVLLLOPLSTAR EQPLVLAIVTAVNQDASNGLTAPNGPQNRVIRKALADAQAQSVGVQVDAVEHGT GTRLDPITAGALLATYQDRAGDEPLWLSVKSNIHTQAAAGMAGVILKMWLRHG LPLRTHVDEPTFDQWSAGAVRLTEETSTGRPRRSAYRSHVSGISTNAHVLEH DPAAGDAGPDPPTSPAAPVPTLSAAEALRAQARLHRRVADGPGRRPLDIA RSLARGALRAVAVLGADETLTSLAALGAGEPDRVVRVTPVSGVPLFTGQ GSGRAGIGRELITETPFARADEVDRLDOLAPLDDLDGRVVDGVLPIEQTGYA LAARFALEVALFVHAGVVDPLVGHSTGELAAHACAGVSLDDACTLVAARGPV QTLTDGMAVAVTSETLAEELIARADVDGRATIAVNGVRSVVSGDADAVERRAR CEERGRKRLPISHAHVSHPMLRLREVAEGLEHPRPRIPMVSNLTGAPVPEA NPVYVROVRETRHDGIRWLGAQVACLELPGGALIMARESLDGPQREVALV PTLQDRPPATPLSALAEHLVHGVADMDTPLAARGATSTDLPTVAFORERYPAT PAAPTAEADRTGRVVRVTPVVDVDAEPCTGTLVAVADGSDGHDGCGARV AALGRAVRLTAGPGLAPALARKLPEDRPVAGVIALPAESAFTWTLAPRSALRGGA RSPASVHPAPVSDTAEAPALPLGLGRVHPVSGGVVDLPGADLARSRLRAVL AAPQDDELALRPGGVPFARKLPAQRPQRQRAWRTGTTLITGDIAGVTELLQYLSP CGARTLLLAHRPQDPLAAGAGGCTTTEWDLGPLETASVIAVHLSDPDPG STPDALTAAGPARAVOURLTVRLBELFDADLDALVLLSTVAGTWGAEDLVHTVA HAALESIAERRLLDGPCTCVGMPWETADTGCASVPGLPLPADLAVSALAELDA DQPCAVDVDSLPFVITGGRFELVADLPVRLAPAPPTAEVLAARAAREHQLL ELVRAQAALVHSGAAIDPVRFDLDFESLAAVFRDRAAEFTGLRLPATLVFDH PTAQAAAHLLASGVQHEDEVAFTADHDDIATVGMACRYPGVGRVRLDVLVH QRLDGVDPPTDRGDLPALRLAPLAKMGGLPDAADFDAAFFGSPREATAMPDQ ORLLRSWQALGAGLDPAALRGSTGVFAGAGSDYGVVLEGAPEADGTYLTTPAA SVLSGLIATLGHAPVAVDTACSSSLVHALHAGLHOGECDLALACVSVIATPG AVDFARQSLADGCKAPAAADGNTNWAEGVGLVVERLSDRNGHRLVATRG AVNDSASGLSPGAQOQVIRQALAAAGLSDVDVAEAGTGTTLGDPLEAQAV LATYGRDRPAGQPLWLSKNTSHSGAAAGGVGVIKMQVALRHGVLPTLHYDETP EVDWSAGTGLLTENREWHPTKPRVGSFAGVGTNNAHVLEQGPETPEAPEEA PEETPGCAPATAPLLPWLISARTPEALRAQARLAARLAERLPAALVGHALLTSR SDFEHAVALGAGHGLTEALRAIAGEPAGTGTGVARPAAGRPVFPVPGOGOMP GMAELDTAPEFAAKABECERALAPYTGFSVELVRSGLTFLERVVVQVPLFAMVS LAALRSYGVVPAAVAGHSGEELACVAGSLDDDAARVVALRAQALRAALRGEGGV SVGIPLAERMLPRWDQVSAVAVGSAVFSAGAPALDDILVAHARDQVRRRID VDYASQVAREIDEITMAGVPQPGQIPFETFTADWDQPTGLDIAVYVRLRE VPLLEPSVQALSGQVDVFEVSPHVPVIGALTEALAGAPVVTGTLRDRDGLKLR VRTALAEVHWGVWTPAFRGHGTTPVELPAYPQRTVWPAAPVAAQDPQDRFE WQAVDQDLPLTAGTGLADEQDRELVPALSARRDRKERAASVSWRIEHLRLE PERFTLPGLVLLVFTGTDTLGTGTDEFTVATGTTGGTADDDVTAALLURE RAEQLVVLAVDPGSDPTTLATLRGALGTEPAGVWSLTDIGSGSPHPRVVPVAGAAI			
Query Match 65.5%; Score 328.2; DB 1; Length 320150; Best Local Similarity 79.5%; Pred. No. 5.8e-27; Matches 403; Conservative 0; Mismatches 58; Indels 6; Gaps 1;			
Qy 1 TCGTCTTCGTGTCGCGTSCACATCGCGGTGACGGCCCTGACGCGCGCGAGTGTCTC 60			
Db 32491 TCGTCTTCGTGTCGCGTSCACCTCGCGCGCGAGGCACTGCGGCGCGCAATGCGAC 32432			
Qy 61 ATGGCGTTCGCGCGCGCGTGTGATGGGACGCTGCGAGAGTTCGTGAGTTCTCG 120			
Db 32431 GTGGCCCTTCGCGCGCGCGTGTGCGGTTCATGGCCACCCCGCGCGCTTCGCG 32372			
Qy 121 CGGCAGCGCGGCTGGCCCCCGACGCGCGCTGCAAGCGTTCGCCGACGCGCGACGCG 180			
Db 32371 CGGAGGGGGGGCTGCGCCCGGAGCGCGGTCAAGAGCGTTCCGCCAGCGGCGACGCG 32312			
Qy 181 ACCGCGTGTTCGAGGGCGTTCGGGCTGCTCTCTGTGTGAGCGCGTTCGAGGCTCAGCGT 240			
Db 32311 ACGGGCTGCGGCGAGGGCGTTCGGCGTACTCTGTGCTGACGCGCTGTCCGAGCGCGCGT 32252			
Qy 241 CGCGGGGACACAGGTCCTCGCGCGTGTTCGGCGGTTCGGCGGTCAAATCTCCACGCGCGTCCG 300			
Db 32251 CGCGGCGATGACGTCGTCTGCGCGTGTTCGGCGGTTCGGCGGTGAACTTCGACGCGTGGTCCG 32192			
Qy 301 AACGGCTTGACGCGCCCGCAACGCGCGTCCAGACGCGCTGATCCGCAAGCACTGGCGC 360			
Db 32191 AACGGTCTGACGCGCGCGCAACGCGTCCGTCGACGAGCGGTGATCCGCGAGCGCTCGT 32132			
Qy 361 GCGCGCGGACTGTCCACATCGGACGCTGACGCGGTGGAGCGCGACCGCGGACGCGAC 420			
Db 32131 TCGCGCGGCTTGTTCGGCGCGCGACGCTGGACATGGTGGAGCGCGACCGCGGACGCTCG 32072			
Qy 421 CTGGGCGACCCGATCGAGCGCGCGGCTGCTGGCGCACCTAGCGGCAGAACCG-----G 474			
Db 32071 CTCGGCGACCCGATCGAGCGCAGGGCTGCTGGCGCACGTAAGTACGACCGCTCCGCGG 32012			
Qy 475 GAAACCGCGCTGTGGCTCGGGTCGGTG 501			
Db 32011 GACCGTTCGCTGTGGCTGGGTTCGGTG 31985			
RESULT 13			
AF453501/c 82746 bp DNA linear BCT 18-JUN-2002			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
misc_feature			
gene			
CDS			



```

/db_xref="GI:21449383"
/translation="ITPGLVVALAATGVLAAPASAAITPTVTSTAITTVEVGRGTG
PAFSAVYDTRSAVTAQAGGIGADHLENGSLVRHSGGGMWANTQVTPPTSAGV
ARWTVRADLGEQASPNVDVVOVEAPVETSAKTHYTAGAGSVAIQSSANVFN
PERGLYHTGDCDKADLADALRYRTTQGITLACVFLAEHRSSPIAQALDRIQV
QFDAQVAGVLKLVFAVYTAESGSDAPKERVLAHDQLALLRANADVAVVQAGFV
GAWGMYTYKFNAGVNSATDOANKRAVDKLTALPADRAVOLRTKFKRLVGGG
STRIGHNDCLASDFTDFTVDPATYFPLAEATRLPMGGETCAVNPRTDCPTAT
AELSRFWYLYNTDQOOLRAWSAQGLAEVAKLYRFTLRDSVFETTTTRGALP
VRVNLTDGYAAAVNORTVNLVLRNTTGAALTEPLTADPTWLGARALIAQDAQIPP
SLATQYQALLLEADPALPGRPEYSIQNTGLMDARTGLNLDKQTIIVS"
1825..2142
/gene="asm36"
/length=318
/gene="asm36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAMS4114.1"
/db_xref="GI:21449382"
/translation="MVTASAVLCARAGCGAPRAGATARWFCSTACKQAAVRLERKA
GSASHEHEDEQLREALIALVARARTLTEATPPWSAQATALVGLARRVVVELAARDAGV
SGG"
complement(2129..3307)
/gene="asm35"
/length=1178
/gene="asm35"
/codon_start=1
/transl_table=11
/product="picyclomycin resistant protein"
/protein_id="AAMS4113.1"
/db_xref="GI:21449381"
/translation="WAVKADSAVGGWLPALLIVALTCTIIEIDMSVPSFPDIARELG
VGGPAVOLVTFLNFGICGLGALCHGSLDREGRPVLGGSAVNLVGLGALAPTIG
LLLRVFCQIGASAAVILVFIIGDVRGPAALRMGLVNAASALMTCAPLGGAL
NRVAGRWGLYACATLALLVFLFPLFETRTARTAITARSVLADRLRLGSRAL
AASITPBLFAAYLVFAASPLVTAAGFGLVGTGFAHLVVVASFAVPSAFATCIIP
LLGGPAGVRMGLGAVVVGFLLLGDGPVAVTGSCVALFCVGPVCPVVFDRMSV
PELAGAASSATMGLRALVLTLLTALAGALSTGGFPFAAVMAGNAVAVLLSALSPR
"
complement(3450..4064)
/gene="asm34"
/length=614
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAMS4112.1"
/db_xref="GI:21449380"
/translation="MARRNPERRAALLDAAVEVLAGEARGLTFRVNDQAGVPAGTA
SNYPAGRAEILTGAERVYQRLDQALPGTDPGDRORSRYTELMHALVDRVSAPTCG
FLALLRLAARNPELREVLTAZRADLRNASHHAASGLPGDGTIVALLWLALNL
ILRLTLPLDLDTHQRDLVTALVDRLLAGQPERAGSPFDESCK"
4151..4723
/gene="asm33"
/length=672
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AAMS4111.1"
/db_xref="GI:21449379"
/translation="WRKLVYFIAASIDQVIAGPGEHDFYPTAPMVEHLRAEYPETT
PTHVRPLVGMPLDTPNDYDVLNMGSGYQPALDAGVGSFYAHLRQTVRSRLSDADP
AVALVRADPVGLVELKREDGLDWLCCGADLQALLPEIDELVVKTYPPVAGGTEPL
FRAGAPARFPDVRAFPHGGVVTYRPA"
complement(4854..5057)
/gene="asm32"
/length=204
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"

```

```

/protein_id="AAMS4110.1"
/db_xref="GI:21449378"
/translation="MWITEQNARIIPARRSSGRSNARRCAVRWPRQPRSATRAPGRS
SANADRSARSGRDRTSLNRSA"
5415..6461
/gene="asm31"
/length=1046
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AAMS4109.1"
/db_xref="GI:21449377"
/translation="WMIGYSADGRRRPVRWGVTEAFTWETDEDIKLAALTRRDAEA
ANADSTRAYLKAGRVRLLTADQVELARVREAGLYAEHLDDLPPAARGSRPVELR
RVRDGQRAKHLLLEANLVLVSTAKRYTGKMLLDVIOEGLNGLVRKVEKDFYAKG
FKSTYATWIRQAITRALADQSRITRILPVMHVVVTKLARLERLLRNREAVTAEE
VAVELDVTPKRVLELQHARTPVSLDQITIGAEADTSIGDLIADENAVAPLEAVSALL
SADLRVLETLTPREAGVVRMYGFPDGVHLTLDVGVGVYGVTRERIRQIESRTWAKL
QHPARSGLGELYD"
complement(6501..9518)
/gene="asm30"
/length=3017
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AAMS4108.1"
/db_xref="GI:21449376"
/translation="MWATGTRIPGPKPLVPLVGNLIDVLTSOLDTVDVDFLDRCHREHGG
I VALTFAGORQVFASFHELVARCMDSQSKAVHPALRQVRDFAGDGLFTAGDRPNW
GKARLLMPAEGPTAMRDPHFAMLDIAEOMLVRRFCFDRHLDVADDWTRLTLTIA
LCAGFARNFYRDRAPHFVDAMVRSLVREAGERLERFGVQFLVGRNQRIFDDIATM
NRIDAVVAARALPAGERPDDLLERMLTCDAPVTGERLSARNVRYQLATELIAGHET
TSGLLSAVHRLLAHPEVLKXADGVLDGRVPAFEDLARDLVIGQVLRTELRLHP
TAPALAPDEPABIGGHAIGAGBPVLMPLTMRDPAVMDPDVDFPERPAPERMDE
IPACAMPFGHGRACITGRPFALQEAATLVALLQRPDLALADPDRLTIKOTLTKP
DSLVRAPRDRPACATVETVTVVQVPAHRHGTPLHVPYGSNGSGEGLARTIAG
DGAARWATSVAPLDDAVRALPASPVVSSINGAPPDNNAAHFVRLTQDQDLRS
VDYLVLGCGNDWSATYQVFTLIDEAAMAAAGARRRGATDARADFFGDWRWYEP
IMPLLSACGVEGEIGFRFRVSDADGDLASAVLENRELVRPDAGSKHLE
LRLPDGTSYRTGDLVLPQNHDPDLVRRARVRLTRAERVTVVTESSAPTLVPVGRAL
RVDELLTRCVDSLAPAGAVARLAERCPPEPRAELAAATGATLLELLEFPSCAVD
LALALELPAETRLYSTSSAAEQRAEVALTVSVGTGTCYLSRVRPDRVAVGAS
PPESFRPADNTVPVLLIAGCTIAPFGFLRAAALGGEGPALLLFCGCGPEHDDL
YAEFPAALGDWLEVDRAVSRHPDGEVRHVQRLMQRDRVRRELVDAGARVYLGGDATR
VGPAAEVYIGRIGFGAGWLDALRAGRYATDVF"
9664..10245
/gene="asm29"
/length=581
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAMS4107.1"
/db_xref="GI:21449375"
/translation="MSAAVRREQLVAALRVMKEDGIAAATTRAICAEADMPHGAFHY
CFRSKQELYTALLTTLGNTSLDAAPPAISPDLPAVDNVRTMLAAFTAFQADPEALV
LDGLVTLAQREPELPELPAWQREYVAKAVAHLDRLAEAGEFDYSVDRALLAEVLVS
LDGVASSWFGSHRDGELAKASLDQFANVMAALTR"
10393..10857
/gene="asm28"
/length=564
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAMS4106.1"
/db_xref="GI:21449374"
/translation="MTDTTTRHRTPVQNAALAVGAVFLLVGIAGFVPLTNTYDGLF
AGHSDAALLGVFKVKSALHNVLHLAGFAGAGIAGLSKAPRGARFLVGGGAIYAVLW
XGVVVDHSTANFVPLNAADNLHLGLAVGMIGLGAALGRNAAHYAARSDD"
complement(10912..11937)

```

CDS	/gene="asm27" complement(10912..11937) /gene="asm27" /codon_start=1 /transl_table=1 /product="glucose 1-dehydrogenase" /protein_id="AA054105.1" /db_xref="GI:21449373" /translation="MPRLAEQVVVVGVGASSIGIRVTAIRFAAGARVVCARNTRAL DGLVEVGRAGRAVATADIAEAAVAVADVVERFGRVDTWNAAGIGVYGRVED TPAGEFDRVRYNYLGHVHGAALPALERAGGGVLIIGVASVILGRSAPLQAPYAASK	
	Query Match 64.9%; Score 325; DB 1; Length 82746; Best Local Similarity 78.0%; Pred. No. 2e-26; Matches 391; Conservative 0; Mismatches 110; Indels 0; Gaps 0;	
QY	1 TCGTCTTCCTCGTGGTGCACATGCGGCGGTGCAGGCCCTCGACGCGCGCGAGTCTCC 60	
Db	61710 TCGTCTCGCTGGTGGCCCTGCACCTGGCGTGGCGTCTGCTCGCGCGAGTCTCG 61651	
QY	61 ATGGCGCTCGCGGCGCGGTGATGTTGATGGGACCGTTCGAGACGTTCTCGAGTTCTCG 120	
Db	61650 CTGGCGCTCGCGGCGCGGTGACGGTTCATGGCCACACCGCGCTGTTCCGCCGACTTCGAC 61591	
QY	121 CGGCAGCGCGGCTGGCCCGACCGCGCTGCAAGCGCTTCGCCGACCGCGCGGACGCG 180	
Db	61590 CGCCAGGCGGGCTCGCGGCGGACGCGCGCTGCAAGCGCTTCGCCGCGGACGCGCG 61531	
QY	181 ACCGCTGGTTCGAGGCGCTCGCGCTGCTCTCGTGGAGCGGTGTCGAGGCTCAGCGT 240	
Db	61530 ACCGCTTCGCGAGGCGCGCGGCTCGTCTGCTGAGCGGTGTCGAGCGCGCGCGC 61471	
QY	241 CGCGGACACAGTCTCGCGCTGCTCGCGGCTCGCGGCTCAACTCCGACGCGCGTGC 300	
Db	61470 AACGGGACGCGCTGTCGCGCTGCTCGCGGCTCAACTCCGACGCGCGTGC 61411	
QY	301 AACGGCTTACGCGCCCGCAACGCGCGTCCGACGCGCTGATCCGCAAGGCACTGGCC 360	
Db	61410 AACGGCTTACGCGCGCGCAACGCGCGCTCCGACGCGGCTGATCCGCGCGCTGGCC 61351	
QY	361 GCGCGCGAGTCTCCATCCGACGCTGACGCGGTGGAGCGCAACGCAACCGGACGACC 420	
Db	61350 GACGCGCGGCTGTCGCGCGCGAGGTGACGCTGTTGAGCGCAACGCAACCGGTA 61291	
QY	421 CTGGCGACCGGATCGAGCGCGGCGCTGCTGCGCACCTACGCGCAGAACCGGGAACG 480	
Db	61290 CTGGGTGACCGATCGAGCGCGGCGGCTGCTGCGCACCTACGCGCGGGACCGGCGG 61231	
QY	481 CCGCTGTGGCTCGGCTCGGTG 501	
Db	61230 CCGGTGCTGTGGCTCGGTG 61210	
RESULT 14		
AF357202	113193 bp DNA linear BCT 04-NOV-2002	
LOCUS	Streptomyces nodosus amphotericin biosynthetic gene cluster, complete sequence.	
DEFINITION	AF357202	
VERSION	AF357202.1 GI:14794889	
KEYWORDS	Streptomyces nodosus	
SOURCE	Streptomyces nodosus	
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 113193)	
AUTHORS	Caffrey, P., Lynch, S., Flood, E., Finnan, S. and O'Liinnk, M.	
TITLE	Amphotericin biosynthesis in Streptomyces nodosus: deductions from analysis of polyketide synthase and late genes	
JOURNAL	Chem. Biol. 8 (7), 713-723 (2001)	
MEDLINE	21344785	
PUBMED	11451671	
REFERENCE	2 (bases 1 to 113193)	
AUTHORS	Caffrey, P., Lynch, S., Flood, E., Finnan, S. and O'Liinnk, M.	
TITLE	Direct Submission	
JOURNAL	Submitted (07-MAR-2001) Industrial Microbiology, University College Dublin, Belfield, Dublin, Ireland	
FEATURES	Location/Qualifiers	
source	1..113193	
gene	/organism="Streptomyces nodosus" /mol_type="genomic DNA" /db_xref="taxon:40318" complement(4..1824) /gene="amphg" complement(4..1824) /gene="amphg" /note="possible amphotericin export protein" /codon_start=1 /transl_table=11 /product="Amphg" /protein_id="AAK73498.1" /db_xref="GI:14794890" /translation="MAPDVPEHEERESEQVRRRLAALLRPHRSVGLALTAGVVGIL LNAFGPLLIGRVTDLIADGVHGGPAPGVDFGALGRLLMLLVLYVVASVFMVON WLYASVRLLIHDLRHAQEKRLPLRYDFDRKPAGETLSRGTDDVNDLQOOLT DLISSVFLIVMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPOVAAQWSAGK LTAHVEMCAGHALVAFDRARAEQREDERNEAVTRAGSGAFSGALPEVMMFVAN LGYVAVVGVKVVNGSILTDGVQAFILYARQFSQPIVEIASVAGRLQSGVASQRV FTLLDAPEOEPEPDRPLAVERVEGFQDFVSFRYSPDPLIEGLSLSEPGSTAVAV GPSGAKGTVANLLMRFYIDSGRIILIDGDTAAWNRDRLSRFGILVLOTWLFKGTI ASNIAYGSPGATRADIVEAARATYADRFRTLSQGYDTVLDDSESGVSGAGEQLITVA RAFLARPVLVDDEATSSVDTTELLIQRAMNTLAPGTSFVIAHRLSTIRDADVIV MKSGRIVEGTHDQLIDAGAYARLHAARADAPADVTVG" complement(1805..3628) /gene="amphh" complement(1805..3628) /gene="amphh" /note="possible amphotericin export protein" /codon_start=1 /transl_table=11 /product="Amphh" /protein_id="AAK73499.1" /db_xref="GI:14794891" /translation="MAPSVRWTFPPQRAAARRQPCGCHSVLLRLMETQLRPVAGSVVA LVGFTALIGLTLLEPFLGALIDEGVVRHSDRIGTIGATMAVAVVAVLQALGAAL LGRVSTLGRDLRSVAVRDLDFSAEIGRFGTSLTSLRSVNDVQVQNLQAGSLGI FVAAPCLGSLVALLAQDVTALILVPMVLVAVCFGLLSRMAALYARLQOITLRI GRLLRRTIGVVRVRSFARDAHEGERTTBELLGLSLGVELIAVMLPSELLMLNL FTLLGVCARRIDSGMOIGALSFLSYLSILMSVNLAFVFLVNPRAVCAERTIT EVLQAGTDVFPASPRMAGPAGQVELVGAEFYPCGAENAVRLDLISLTPGSRVAVL GSGSGKTLILHLRLVDVTAGEVRIIGGTDVRELPDSVMAAVGVVPQRPYLFATV ASNMRFPDADDEELWEVRIAQHEGFVRLGLDTEIAQGGTIVSGGQRIALAR ALIRRPAYLFDSDSALDQSTEALRLKALVPYEGATVITVAQRVSVASVRDADRIVLL DOGGIAATGTHALLRDSPTYREIALSQRTEETHAGAGRS" 3840..4874 /gene="amphdIII" 3840..4874 /genes="amphdIII" /notes="mycosamine biosynthesis protein" /codon_start=1 /transl_table=11 /product="AmphdIII" /protein_id="AAK73500.1" /db_xref="GI:14794892" /translation="MPKRALITGIGDQGSYLAHLSSQGVQVWGLIRGOANPRKFRV SLRSLISFVGDGLQGSLSVADVKVQDEVYNLGAISFVPMWSQOAEILVTRVNGM VRLVLEAIRMSVGLSMRSTAGTGOIRFYQASSEMFMKVAETPQRETLTFHRSPYG AAKAYGHFTRNRYRESFGMYAVSGMLPNHSPERGQEFVTRKISLAVARIKLIQDQKL ALGNDAVDWGIVGAPVYRAMHMLQDQDDDDTVIGTGEMHTVDRDAVFAFEVGLDM KQVYVDPDLVRPAEVEVLCASSKAQALQIGKPSVDQFELMMVMVDADLASVSRQNE LDLLILAHAS" 5042..33574 /gene="amphI" 5042..33574 /genes="amphI" /note="polyketide synthase multienzyme polypeptide housing	





Thu Jun 17 08:25:05 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:08:13 ; Search time 285 Seconds  
(without alignments)  
7467.885 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tcgttcgctggctgcgct.....cgctgctggctggctggctg 501

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	501	100.0	53789	2 AAV21187	Amycolato
2	334.6	66.8	5676	2 AAV21186	Amycolato
3	332	66.3	125401	4 AAD17186	Streptomy
4	328.2	65.5	20394	4 AAF24892	Pimaricin
5	325	64.9	9975	7 AAL61173	Actinosyn
6	325	64.9	82746	7 AAL61224	Actinosyn
7	320.8	64.0	113193	7 AAD54645	Streptomy
8	320.2	63.9	9222	7 AAL61171	Actinosyn
9	320	63.9	14055	7 AAL61170	Actinosyn
10	317.6	63.4	65140	4 AAD17184	Streptomy
11	314.4	62.8	77536	3 AAL14651	Nucleotid
12	307.4	61.4	15872	3 AAZ87283	S. venezu
13	307.4	61.4	43280	2 AAT80413	Tylactone
14	307.4	61.4	103599	4 ABX04971	S. cinnam
15	305.8	61.0	15872	2 AAT68715	Streptomy
16	305.8	61.0	38506	3 AAZ75633	Nucleotid
17	305.8	61.0	38506	3 AAZ56001	Recombina
18	305.8	61.0	38506	7 ABS56090	S. venezu
19	305.8	61.0	38506	7 ADA09418	Cosmid pk
20	304.2	60.7	4674	3 AAL14666	Nucleotid
21	304.2	60.7	4725	3 AAL14665	Nucleotid
22	304.2	60.7	4737	3 AAL14668	Nucleotid
23	304.2	60.7	4767	3 AAL14667	Nucleotid

24	304.2	60.7	4818	3 AAL14669	Nucleotid
25	304.2	60.7	36778	3 AAZ87318	S. venezu
26	304.2	60.7	37948	3 AAZ87285	S. venezu
27	303.8	60.6	13842	3 AAZ87297	S. venezu
28	303.2	60.5	77536	3 AAL14651	Nucleotid
29	302.6	60.4	6210	7 ACC42734	Geldanamy
30	300	59.9	103599	4 ABX04971	S. cinnam
31	297	59.3	30690	3 AAZ92301	S. cinnam
32	297	59.3	30690	4 AAZ92301	Streptomy
33	294.2	58.7	1419	6 AAS18433	Contig 14
34	294.2	58.7	27541	4 AAD17185	Streptomy
35	293	58.5	4466	3 AAL14663	Nucleotid
36	293	58.5	4478	3 AAL14661	Nucleotid
37	293	58.5	4547	3 AAL14664	Nucleotid
38	292.8	58.5	4571	3 AAL14662	Nucleotid
39	292.8	58.4	44377	2 AAT80414	Platenoli
40	292.8	58.4	44377	2 AAT78508	Platenoli
41	290.6	58.0	6360	7 ACC42735	Geldanamy
42	288.8	57.6	110000	4 AAI99682_18	Continuation (19 o
43	288.8	57.6	110000	4 AAI99683_18	Continuation (19 o
44	288.2	57.5	29879	2 AAQ46806	eryA regi
45	287.8	57.4	5778	7 AAD55816	Micromono

ALIGNMENTS

RESULT 1  
AAV21187  
ID AAV21187 standard; DNA; 53789 BP.  
XX  
AC AAV21187;  
XX  
DT 24-JUL-1998 (first entry)  
DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.  
XX  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
XX polyketide synthase; actinomycete; ansamycin; ds.  
XX  
OS Amycolatopsis mediterranei.  
XX  
FH Key Location/Qualifiers  
FT CDS 1825..15543  
FT /tag= a  
FT /label= ORF A  
FT /product= "polyketide synthase"  
FT 15550..30759  
FT /tag= b  
FT /label= ORF B  
FT /product= "polyketide synthase"  
FT 30895..36060  
FT /tag= c  
FT /label= ORF C  
FT /product= "polyketide synthase"  
FT 36259..41325  
FT /tag= d  
FT /label= ORF D  
FT /product= "polyketide synthase"  
FT 41373..51614  
FT /tag= e  
FT /label= ORF E  
FT /product= "polyketide synthase"  
FT 51713..5293  
FT /tag= f  
FT /label= ORF F  
FT /product= "polyketide synthase"

WO9807868-A1.

26-FEB-1998.

18-AUG-1997; 97WO-EF004495.



Db 3897 TCCTCGTGGTGGCGATGCACCTGGCGCGGACGCGCTGGCGGACGCGGAGTGTCTCG 3956  
QY 61 ATGGCGCTGCGGGCGCGGTGATGGTGGGACACGTCGAGACGCTTCGTGAGTTCCTCG 120  
Db 3957 ATGGCGCTGCGGGCGCGGTGATGGTGGGACACGTCGAGACGCTTCGTGAGTTCCTCG 4016  
QY 121 CGGACGCGGGGCTGGCGCGGACGCGGCTGCAAGCGCTTCGCCACGCGCGGACGCGC 180  
Db 4017 AAGCAGCGGGCTGGCGCGGACGCGGCTGCAAGCGCTTCGCCACGCGCGGACGCGC 4076  
QY 181 ACCGGTGTTCGAGGCGGCTGGCGGCTGCTCTCTGTTGAGGCGGTTCGCGAGGCTCAGCGT 240  
Db 4077 ACGGCTGTGGCGGACGCGGCTGGCGGCTGCTCTCTGTTGAGGCGGTTCGCGAGGCTCAGCGT 4136  
QY 241 CGGCGGACGCGGCTTCGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 300  
Db 4137 CGCGGCGACCGGGTGTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCTTC 4196  
QY 301 AAGCGCTTACGCGCGCGCGGACGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 360  
Db 4197 AAGCGCTTACGCGCGCGCGGACGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 4256  
QY 361 GCGCGCGGACTGTTCACATCGGACGCTGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 420  
Db 4257 GCGCGCGGCTTCGAAACGCTGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 4316  
QY 421 CTGGCGGACCGGCTGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 474  
Db 4317 CTGGCGGACCGGCTGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 4376  
QY 475 GAAACGCGGCTGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 501  
Db 4377 GAGACGCGGCTGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGGCGTGG 4403

## RESULT 3

AA017186  
ID AAD17186 standard; DNA; 125401 BP.  
XX  
AC AAD17186;  
XX  
XX  
DT 29-NOV-2001 (first entry)  
DE Streptomyces noursei nystatin PKS gene cluster DNA.  
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; ds.  
OS Streptomyces noursei.  
FH Location/Qualifiers  
FT 6337..34771  
FT /\*tag= a  
FT /product= "NysI complete protein"  
FT 34792..51099  
FT /\*tag= b  
FT /product= "NysJ protein"  
FT 51155..57355  
FT /\*tag= c  
FT /product= "NysK protein"  
FT 57503..58687  
FT /\*tag= d  
FT /product= "NysL protein"  
FT complement(58786..58980)  
FT /\*tag= e  
FT /product= "NysM protein"  
FT /note= "CDS does not include start codon"  
FT complement(59045..60241)  
FT /\*tag= f  
FT /product= "NysN protein"  
FT /note= "CDS does not include start codon"  
FT complement(60238..61296)  
FT /\*tag= g  
FT

FT CDS  
FT /product= "NysD2 complete protein"  
FT 120628..121308  
FT /\*tag= h  
FT /product= "NysR4 (long) protein"  
XX  
PN W0200159126-A2.  
XX 16-AUG-2001.  
XX 08-FEB-2001; 2001WO-GB0000509.  
XX 08-FEB-2000; 2000GB-00002840.  
XX 10-APR-2000; 2000GB-00008786.  
XX 14-APR-2000; 2000GB-00009387.  
XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPSELIGE.  
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.  
PA (ALPH-) ALPHARMA AS.  
PA (SINV-) SINVENT AS.  
PA (DZIE/) DZIELEWSKA H.  
PA (ZOTC/) ZOTCHEV S B.  
PA (SEKU/) SEKUROVA O N.  
PA (FJAE/) FJAEVIRK E.  
PA (BRAU/) BRAUTASET T.  
PA (STRO/) STROM A R.  
PA (VALL/) VALLA S.  
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
XX WPI; 2001-557614/62.  
DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,  
DR AAE10149, AAE10150.  
XX New nystatin polyketide synthase polynucleotides and polypeptides, useful  
FT as antibiotics and antifungals.  
PS Claim 1; Page 188-254; 266pp; English.  
XX  
XX The present invention relates to the cloning and sequencing of the gene  
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
CC involved in the biosynthesis of the macrolide antibiotic nystatin. The  
CC nystatin PKS is useful as antifungal antibiotics. The present sequence is  
CC a Streptomyces noursei nystatin PKS gene cluster DNA  
XX  
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;  
Query Match 66.3%; Score 332; DB 4; Length 125401;  
Best Local Similarity 79.0%; Pred. No. 1.8e-43;  
Matches 395; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 1 TCCTCTTCGCTGGTGGCGCTGCACATGGCGCTGCAGGCGCTGCAGGCGGCGGAGTGTCTCC 60  
Db 35431 TCCTCTTCGCTGGTGGCGCTGCACATGGCGCTGCAGGCGCTGCAGGCGGCGGAGTGTCTCC 35490  
QY 61 ATGGCGCTTCGCGGCGGCGCTGATGGTGGTGGCGGCTGCAGGCGCTTCGTGAGTTCCTCG 120  
Db 35491 ATGGCGCTTCGCGGCGGCGCTGATGGTGGTGGCGGCTGCAGGCGCTTCGTGAGTTCCTCG 35550  
QY 121 CGGACGCGGGGCTGGCGCGGACGCGGCTGCAAGCGCTTCGCCACGCGCGGACGCGC 180  
Db 35551 CGGACGCGGGGCTGGCGCGGACGCGGCTGCAAGCGCTTCGCCACGCGCGGACGCGC 35610  
QY 181 ACCGGTGTTCGAGGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGT 240  
Db 35611 ACCGGTGTTCGAGGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGT 35670  
QY 241 CGCGGCGACGAGGCTTCGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGT 300  
Db 35671 AAGCGCGACCGGCTTCGCGGCTGGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGT 35730  
QY 301 AACGGCTTACGCGGCGGCGGCTTCGCCACGCGGCTGGCGGCTGGCGGCTGCAAGCGCTTCGCCACGCGCGT 360





CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
 CC useful as a potent antitumor agent. The present sequence is A. pretiosum  
 CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
 XX  
 SQ Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;  
 Query Match 64.9%; Score 325; DB 7; Length 9975;  
 Best Local Similarity 78.0%; Pred. No. 3.4e-42;  
 Matches 391; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
 1 TCGTCTTCCTGCTGCGTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCC 60  
 Db TCGTCTTCCTGCTGCGTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCG 5550  
 61 ATGGCGCTCGCGGCGGTGATGATGATGGGACCGGTGAGACGCTTCGTGAGTTCCTCG 120  
 Db ATGGCGCTCGCGGCGGTGATGATGATGGGACCGGTGAGACGCTTCGTGAGTTCCTCG 5551  
 121 CGGACGCGGGGTGCGCCCGGACGCGCTGCAAGGCGTTCGCGACGCGCGGACGCGC 180  
 Db CGGACGCGGGGTGCGCCCGGACGCGCTGCAAGGCGTTCGCGACGCGCGGACGCGC 5611  
 181 ACCGGCTGCTCGAGGCGGTGCGGCTGCTCTGCTGAGCGGCTGCTCCGAGGCTCAGCGT 240  
 Db ACCGGCTGCTCGAGGCGGTGCGGCTGCTCTGCTGAGCGGCTGCTCCGAGGCTCAGCGT 5671  
 241 CGGCGGACAGGTCTCTGCGGCTGCTCGCGGCTGCGGCGTCAACTCCGACGCGGCTCG 300  
 Db CGGCGGACAGGTCTCTGCGGCTGCTCGCGGCTGCGGCGTCAACTCCGACGCGGCTCG 5731  
 301 AAGCGGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360  
 Db AAGCGGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5791  
 361 GCGCGCGGCTGCTCCATCGGACGCTGCGAGCGGTGCGAGCGCGCACCGGCGACGACC 420  
 Db GCGCGCGGCTGCTCCATCGGACGCTGCGAGCGGTGCGAGCGCGCACCGGCGACGACC 5851  
 421 CTGGGCGACCGGATCGAGCGGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCG 480  
 Db CTGGGCGACCGGATCGAGCGGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 5911  
 481 CGGCTGTGCTCGGGTCTGGTG 501  
 Db CGGCTGTGCTCGGGTCTGGTG 5971

## RESULT 6

AAL61224/c

ID AAL61224 standard; DNA; 82746 BP.

XX

XX AAL61224;

XX

22-SEP-2003 (first entry)

XX

Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.

DE

XX Maytansinoid; ansamitocin; antitumor; gene; ds.

XX

OS Actinosynnema pretiosum.

XX

XX WO2003045312-A2.

XX

XX 05-JUN-2003.

XX

XX 21-NOV-2002; 2002WO-US037547.

XX

XX 21-NOV-2001; 2001US-0332158P.

XX

XX (UNIW ) UNIV WASHINGTON.

XX

XX Floss HG, Yu T, Leistner E;

XX

XX

XX

XX

XX

DR WPI; 2003-493374/46.

XX Novel maytansinoid produced by bacterial host cell transformed with

PT expression vector comprising open reading frame from ansamitocin gene

PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.

XX

PS Claim 7; Page 105-152; 160pp; English.

XX

XX The invention relates to maytansinoid produced by bacterial host cell

CC transformed with expression vector comprising open reading frame from

CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is

CC useful as a potent antitumor agent. The present sequence is A. pretiosum

CC ansamitocin ansamitocin biosynthetic gene cluster I

XX

SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;

Query Match 64.9%; Score 325; DB 7; Length 82746;

Best Local Similarity 78.0%; Pred. No. 2.4e-42;

Matches 391; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

1 TCGTCTTCCTGCTGCGTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCC 60

Db TCGTCTTCCTGCTGCGTGCACATGGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCTCG 61710

61 ATGGCGCTCGCGGCGGTGATGATGATGGGACCGGTGAGACGCTTCGTGAGTTCCTCG 120

Db ATGGCGCTCGCGGCGGTGATGATGATGGGACCGGTGAGACGCTTCGTGAGTTCCTCG 61650

121 CGGACGCGGGGTGCGCCCGGACGCGCTGCAAGGCGTTCGCGACGCGCGGACGCGC 180

Db CGGACGCGGGGTGCGCCCGGACGCGCTGCAAGGCGTTCGCGACGCGCGGACGCGC 61590

181 ACCGGCTGCTCGAGGCGGTGCGGCTGCTCTGCTGAGCGGCTGCTCCGAGGCTCAGCGT 240

Db ACCGGCTGCTCGAGGCGGTGCGGCTGCTCTGCTGAGCGGCTGCTCCGAGGCTCAGCGT 61530

241 CGCGCGGACCGAGTCTCTGCGGCTGCTCGCGGCTCGGCGGTCAACTCCGACGCGGCTCG 300

Db CGCGCGGACCGAGTCTCTGCGGCTGCTCGCGGCTCGGCGGTCAACTCCGACGCGGCTCG 61470

301 AAGCGGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360

Db AAGCGGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61410

361 GCGCGCGGCTGCTCCATCGGACGCTGCGAGCGGTGCGAGCGCGCACCGGCGACGACC 420

Db GCGCGCGGCTGCTCCATCGGACGCTGCGAGCGGTGCGAGCGCGCACCGGCGACGACC 61350

421 CTGGGCGACCGGATCGAGCGGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 480

Db CTGGGCGACCGGATCGAGCGGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGG 61290

481 CGGCTGTGCTCGGGTCTGGTG 501

Db CGGCTGTGCTCGGGTCTGGTG 61230

## RESULT 7

AAD54645

ID AAD54645 standard; DNA; 113193 BP.

XX

XX AAD54645;

XX

XX 26-JUN-2003 (first entry)

XX

XX Streptomyces nodosus amphoterin (amph) biosynthetic gene cluster.

XX

XX Polyene; antibiotic; amphoterin; amph; polyketide; enzyme; gene; ds.

XX

XX Streptomyces nodosus.

XX

XX Key Location/Qualifiers

XX

XX complement (4. .1824)

XX

XX

```
FT FT /tag= a
FT FT /product= "ABC transporter encoded by S. nodosus amphG
FT FT gene"
FT CDS complement(1805..3628)
FT FT /tag= b
FT FT /product= "ABC transporter encoded by S. nodosus amphH
FT FT gene"
FT CDS 3840..4874
FT FT /tag= c
FT FT /product= "GDP-mannose dehydratase encoded by S. nodosus
FT FT amphiIII gene"
FT CDS 5042..33574
FT FT /tag= d
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT FT nodosus amphiK gene"
FT CDS 33584..50518
FT FT /tag= e
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 15, 16 and 17 encoded by S. nodosus
FT FT amphiJ gene"
FT CDS 50571..56675
FT FT /tag= f
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 18 and thioesterase encoded by S.
FT FT nodosus amphiK gene"
FT CDS 56829..58019
FT FT /tag= g
FT FT /product= "Cytochrome P450 encoded by S. nodosus amphiL
FT FT gene"
FT CDS 58139..58648
FT FT /tag= h
FT FT /product= "ORF1, hypothetical protein"
FT FT complement(58756..59610)
FT FT /tag= i
FT FT /product= "ORF2, hypothetical protein"
FT FT 59869..61470
FT CDS /tag= j
FT FT /product= "ORF3, hypothetical protein"
FT FT /transl_except= (pos:59869..59871, aa:Met)
FT CDS complement(61798..61995)
FT FT /tag= k
FT FT /product= "Ferredoxin encoded by S. nodosus amphiM gene"
FT CDS complement(62051..63250)
FT FT /tag= l
FT FT /product= "Cytochrome P450 encoded by S. nodosus amphiN
FT FT gene"
FT CDS complement(63250..64308)
FT FT /tag= m
FT FT /product= "NDP-sugar aminotransferase encoded by S.
FT FT nodosus amphiII gene"
FT CDS complement(64324..65775)
FT FT /tag= n
FT FT /product= "Glycosyl transferase encoded by S. nodosus
FT FT amphiD gene"
FT FT /transl_except= (pos:65773..65775, aa:Met)
FT CDS 66081..70319
FT FT /tag= o
FT FT /product= "Polyketide synthase multienzyme housing
FT FT loading module encoded by S. nodosus amphiA gene"
FT FT 70366..79938
FT CDS /tag= p
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 1 and 2 encoded by S. nodosus by amphiB
FT FT gene"
FT FT 79956..112709
FT CDS /tag= q
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT FT nodosus by amphiC gene"
FT XX
```

WO200297082-A2.

```
PD 05-DEC-2002.
XX
XX 27-MAY-2002; 2002WO-IE000071.
XX
XX 31-MAY-2001; 2001IE-00000527.
XX
XX (UYDU-) UNIV COLLEGE DUBLIN.
XX
XX Caffrey JP;
XX
XX WPI; 2003-201271/19.
DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
DR AAE36129, AAE36130, AAE36131, AAE36132.
XX
XX Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX for preparing amphotericin derivative or analog antibiotic agent with
XX altered properties, in biosynthesis of polyketide other than
XX amphotericin.
XX
XX Claim 1; Page 52-114; 276pp; English.
XX
XX The invention relates to the gene cluster encoding the polypeptides
XX responsible for the biosynthesis of the polyene antibiotic amphotericin
XX (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX useful for preparing amphotericin derivatives or analogue antibiotic
XX agents with altered properties and in the biosynthesis of polyketides
XX other than amphotericin. amphiIII, amphiII or amphiI mutants are useful
XX for producing amphotericin derivatives glycosylated with alternative
XX sugars; amphiIII or amphiII gene sequences are useful in engineered
XX biosynthesis of peramanyl-amphoteronolide B; amphiIII or amphiII and
XX amphiN gene sequences are useful in the engineered biosynthesis of
XX peramanyl-16-desacboxyl-16-methyl amphoteronolide B; amphiIII, amphiII
XX and amphiD gene sequences are useful for preparing polypeptides capable
XX of addition of mycosamine to a polyketide other than amphoteronolide A or
XX B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX The present sequence is S. nodosus amphi biosynthetic gene cluster
XX
XX Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX
XX Query Match 64.0%; Score 320.8; DB 7; Length 113193;
XX Best Local Similarity 77.6%; Pred. No. 1e-41;
XX Matches 388; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 1 TGTCTTCCTGCTGCTGCTGCACATGGCGGTGAGGCGCTTCGAGACCGTTCGTTCGAGTTCTCG 60
Db 80568 TCTCTCTCTGCTGCTTCCTCTGACCTGGCGCGGAGGCGCTGCGCGGCGAGTCTCC 80627
QY 61 ATGGCGCTCGCGGCGGCGTGTGATGGGACCGTTCGAGACCGTTCGTTCGAGTTCTCG 120
Db 80628 CTGGCCCTTGGCGGCGGCGTCCGTGATGGCGACCGCCGATCATCGGTTTCACTCC 80687
QY 121 CGGCAGCGGGGCTGCCCCCGACCGCCCTGCAGGCGCTTCGCGAGCGGCGGACGCG 180
Db 80688 TCGCAGAGCGGACTGGCCCGCGACGCGCGCTGCAAGCGTTCTTCGACGCGCGGCGG 80747
QY 181 ACCGCTGTCTCGAGGCGCTCGGCGTCTGCTGTGGAGCGGCTGTCGAGGCTCAGCGT 240
Db 80748 ACCGCTGTGGCGGCGAGGCGCGCGCTGTTCTGATGGGCTGTCGAGCGCCGAGCGC 80807
QY 241 CGCGGCGCACAGGCTCTCGCGGTGCTCGCGGCTCGCGGCTCAACTCCGAGCGCGTGG 300
Db 80808 GAGGCGCGCAAGGCTCTCGCGGTACTCGCGGCTCGCGGCTCAACCGAGGCGCGCTCC 80867
QY 301 AACGCTTACGCGCCCGGAAACGCGCCCGTCCAGCAGCGGTGATTCGCGAGGCACTGGCC 360
Db 80868 AACGCGCTGACCGCGCCCGAAACGCGCCCGTCCCAACAGCGGCTCATCGGCGGCGG 80927
QY 361 GCGCGCGGACTGTTCACATCGGACGTCGACGCGGCTGGAGCGCGACCGCGACCGGACG 420
Db 80928 CACGCGGCTGTGAGCGCGAGCGATGTCGCGCTGAGCGCGATGCGGCGCGGCGCGG 80987
QY 421 CTGGCGGCGACCGCGATCGAGGCGGCGCTGCTGGCCACCTTACGCGCAGAACCGGGAACG 480
```

```

Db      80988 CTCGGTACCCGATCGAGGCAAGGCCCTGCTGGCCACCTACGGCCAGAACCGCGAGCGG 81047
QY      481 CCGCTGTGGCTCGGGTCCGT 500
Db      81048 CCGCTGTGCTCGGCTCCGT 81067

RESULT 8
AAL61171
ID AAL61171 standard; DNA; 9222 BP.
XX
AC AAL61171;
XX
XX 22-SEP-2003 (first entry)
XX
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #2.
DE
DE Actinosynnema pretiosum; antitumour; polyketide synthase; enzyme; PKS;
XX
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
XX
XX gene; ds.
XX
XX Actinosynnema pretiosum.
XX
XX WO2003045312-A2.
XX
XX 05-JUN-2003.
XX
XX 21-NOV-2002; 2002WO-US037547.
XX
XX 21-NOV-2001; 2001US-0332158P.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Floss HG, Yu T, Leistner E;
XX
XX WPI; 2003-493374/46.
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Disclosure; Page 53-58; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin gene cluster I polyketide synthase (PKS) gene
XX
XX Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;
XX
XX Query Match 63.9%; Score 320.2; DB 7; Length 9222;
XX Best Local Similarity 78.5%; Pred. No. 1.9e-41;
XX Matches 398; Conservative 0; Mismatches 103; Indels 6; Gaps 1;
QY      1 TCGTCTTCGCTGTCGCTGACATGCGGTGAGGCCCTCGACGCGGCGAGTGTCTCC 60
Db      604 TCGTCTTCGCTGTCGCTGACATGCGGTGAGGCCCTCGACGCGGCGAGTGTCTCG 663
QY      61 ATGGCGCTCGCGGGCGGCTGATGCTGATGCGACGCTCGAGAGTCTCG 120
Db      664 ATGGCGCTCGCGGGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 723
QY      121 CGGACGCGCGGCTGCGCCCGCCGACGCGCGCTGCAAGAGCGTTTCGCCAGCGCGCGACGCGC 180
Db      724 CGGACGCGCGCTGCGCCCGCCGACGCGCGCTGCAAGAGCGTTTCGCCAGCGCGACGCGC 783
QY      181 ACCGCTGTGTCGAGGCGCTCGGCTGCTCCCTGTTGTTGAGCGGCTGTCCGAGGCTCAGGCT 240
Db      784 ACCGCTGTGTTGAGGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 843
QY      241 CGCGGGCACAGTCTTCGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 300
```

```

Db      844 GACGGGCGCGGGTGTGGCGGTGTGGCGGTGTGGCGGTGTGGCGGTGTGGCGGTGTGGCG 903
QY      301 AACGGCTTACGCGCGCCCGAAGCGCCCGTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC 360
Db      904 AACGGCTTACGCGCGCCCGAAGCGCCCGTCCAGCAGCGCGTGTATCCGCGCGCGCTGGCC 963
QY      361 GCGCGCGGACTGTGTCACATCGGACGTCGACGCGGTGAGCGCGCACCGGACCGGACGACC 420
Db      964 GACGGCGGCTGTGGCGCGGACGTCGACCTGCTGAGGCGCGCACCGGACCGGACCGGTG 1023
QY      421 CTGGGCGACCCGATCGAGCGCGGCGCTGCTGCGCCACCTACGCGCCAGAACCC-----GG 474
Db      1024 CTGGGCGACCCGATCGAGCGCGGCGCGCTGCTGAGACCTTACGCGCGCGGTTCGCGCGGC 1083
QY      475 GAAACGCGCTGTGGCTCGGGTCCGGT 501
Db      1084 GCGGGCGCGCTGTGGCTCGGGTCCGGT 1110

RESULT 9
AAL61170
ID AAL61170 standard; DNA; 14055 BP.
XX
AC AAL61170;
XX
XX 22-SEP-2003 (first entry)
XX
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #1.
DE
DE Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
XX
XX gene; ds.
XX
XX Actinosynnema pretiosum.
XX
XX WO2003045312-A2.
XX
XX 05-JUN-2003.
XX
XX 21-NOV-2002; 2002WO-US037547.
XX
XX 21-NOV-2001; 2001US-0332158P.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Floss HG, Yu T, Leistner E;
XX
XX WPI; 2003-493374/46.
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Disclosure; Page 45-53; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin gene cluster I polyketide synthase (PKS) gene
XX
XX Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;
XX
XX Query Match 63.9%; Score 320; DB 7; Length 14055;
XX Best Local Similarity 78.6%; Pred. No. 1.9e-41;
XX Matches 396; Conservative 0; Mismatches 105; Indels 3; Gaps 1;
QY      1 TCGTCTTCGCTGTCGCTGACATGCGGTGAGGCCCTCGACGCGGCGAGTGTCTCC 60
Db      3550 TCGTCTTCGCTGTCGCTGACATGCGGTGAGGCCCTCGACGCGGCGAGTGTCTCG 3609
QY      61 ATGGCGCTCGCGGGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 120
Db      3610 ATGGCGCTCGCGGGCGGCGGCGAGCGTCTATGGCCACCGCGCGGCTTCGCTGAGTTCGCC 3669
```



[illegible]

```
FT misc_feature complement(31018. .32185)
FT /tag= ag
FT /note= "AT3"
FT misc_feature complement(32185. .33439)
FT /tag= ah
FT /note= "K33"
FT misc_feature complement(33505. .33715)
FT /tag= ai
FT /note= "ACP2"
FT misc_feature complement(33823. .34480)
FT /tag= aj
FT /note= "KR2"
FT misc_feature complement(34506. .35749)
FT /tag= ak
FT misc_feature complement(35749. .37144)
FT /tag= am
FT /note= "AT2"
FT misc_feature complement(37145. .38296)
FT /tag= an
FT /note= "KS2"
FT misc_feature complement(38371. .38581)
FT /tag= ao
FT /note= "ACP1"
FT misc_feature complement(38677. .39307)
FT /tag= ap
FT /note= "KR1"
FT misc_feature complement(39442. .40609)
FT /tag= aq
FT /note= "DH1"
FT misc_feature complement(40609. .41842)
FT /tag= ar
FT /note= "AT1"
FT misc_feature complement(41842. .43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT misc_feature complement(43144. .43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT misc_feature complement(43777. .44629)
FT /tag= au
FT /note= "ER of loading domain"
FT misc_feature complement(44974. .46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT CDS 46754. 47788
FT /tag= aw
FT /note= "fkb gene"
FT CDS 47785. 52272
FT /tag= ax
FT /note= "fkb gene"
FT CDS 52275. 71465
FT /tag= ay
FT /note= "fkb gene"
FT misc_feature 52362. 53576
FT /tag= az
FT /note= "KS7"
FT misc_feature 53577. 54716
FT /tag= bb
FT /note= "AT7"
FT misc_feature 54717. 55871
FT /tag= ba
FT /note= "DH7"
FT misc_feature 56019. 56819
FT /tag= bc
FT /note= "KR7"
FT misc_feature 56943. 57575
FT /tag= bd
FT /note= "KR7"
FT misc_feature 57710. 57920
FT /tag= be
FT /note= "ACP7"
FT misc_feature 57990. 59243
FT /tag= bf
```

```
FT misc_feature /note= "KS8"
FT 59244. 60398
FT /tag= bg
FT /note= "AT8"
FT 60399. 61412
FT /tag= bh
FT /note= "DH8 (inactive)"

Query Match 62.8%; Score 314.4; DB 3; Length 77536;
Best Local Similarity 76.8%; Pred. No. le-40;
Matches 384; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 1 TCGTCTTCGTCGCGCTGCACATGCGCGTGCAGCCCTGCAGCGCGGAGTGTCTCC 60
Db TCGTCGTCGTCGCGCTGCACATGCGCGTGCAGCGCGTGCAGCGCGGAGTGTCTCC 42533
Qy 61 ATGGCGCTCGCGCGCGCGTGCATGATGATGATGATGATGATGATGATGATGATG 120
Db CTGGCGCTCGTCGCGCGTGCATGATGATGATGATGATGATGATGATGATGATGAT 42473
Qy 121 CGGCAGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAG 180
Db CGGCAGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCG 42413
Qy 181 ACCGCGCTCGTCGAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCG 240
Db ACCGCGCTCGTCGAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCG 42353
Qy 241 CGCGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCG 300
Db AACGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCG 42353
Qy 301 AACGCGTTCGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCG 360
Db AACGCGTTCGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGC 42233
Qy 361 GCCTCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCG 420
Db AACGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCG 42173
Qy 421 CTGGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCG 480
Db CTGGCGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCGTGCAGCGCG 42113
Qy 481 CGCTGTGCTCGCGTGCAGT 500
Db 42112 CGCTGTGCTCGCGTGCAGT 42093

RESULT 12
AAZ87283
ID AAZ87283 standard; DNA; 15872 BP.
XX AAZ87283;
AC AAZ87283;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW Desamine biosynthesis; macrolide; polyketide; polymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae; ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT /tag= a
FT /product= "vep ORF 1 amino acid sequence #1 (AAZ77177)"
FT 14056. 14151
FT CDS
```

FT		/*tag= b	
FT	CDS	/product= "vcp ORF 1 amino acid sequence #3 (AAV77199)"	
FT		/length= .15827	
FT		/*tag= c	
FT		/product= "vcp ORF 1 amino acid sequence #2 (AAV77178)"	
XX			
XX	WO20000620-A2.		
PN			
XX	06-JAN-2000.		
XX			
XX	25-JUN-1999;	99WO-USO14398.	
PF			
XX	26-JUN-1998;	98US-00105537.	
XX	(MINU ) UNIV MINNESOTA.		
PA			
XX	Sherman DH, Liu H, Xue Y, Zhao L;		
PI	WPI; 2000-160679/14.		
XX	P-PSDB; AAY77177, AAY77178, AAY77199.		
DR			
XX			
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.		
XX	synthesis of methymycin and pikromycin.		
XX			
XX	Example 3; Fig 23; 438pp; English.		
PPS			
XX	The invention relates to an isolated and purified nucleic acid segment		
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its		
CC	biologically active variant, where the nucleic acid sequence is not		
CC	derived from the eryc gene cluster of Saccharopolyspora erythraea or		
CC	Streptomyces antibioticus. The invention also relates to a macrolide		
CC	biosynthetic gene cluster, or fragments thereof. The macrolide		
CC	biosynthetic gene cluster encodes proteins which synthesize methymycin,		
CC	pikromycin, neomethymycin, narbomycin or a combination of these		
CC	compounds. Recombinant or augmented cells comprising the desosamine		
CC	and/or macrolide biosynthetic gene clusters are useful for the production		
CC	of biologically active macrolides. The macrolide biosynthetic proteins		
CC	are useful for synthesis of methymycin, pikromycin, neomethymycin and		
CC	narbomycin. The alternative termination of polyketide synthesis may be		
CC	useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)		
CC	monomers. The compounds produced by the recombinant host cells are useful		
CC	as biopolymers, e.g., in packaging or biomedical applications, to		
CC	engineer PHA monomer synthases or to prepare biologically active agents,		
CC	such as chemotherapeutics, immunosuppressants, agents to treat asthma,		
CC	chronic obstructive pulmonary disease as well as other diseases involving		
CC	respiratory inflammation, cholesterol-lowering agents or macrolide-based		
CC	antibiotics which are active against a variety of organisms, e.g.,		
CC	bacteria, including multi-drug resistant pneumococci and other		
CC	respiratory pathogens, as well as viral parasitic pathogens, or as crop		
CC	protection agents (e.g., fungicides or insecticides) via expression of		
CC	polyketides in plants. The present sequence represents a Streptomyces		
CC	vanderuelae ATCC 15439 DNA sequence, designated vep ORF 1 in the		
CC	specification, which actually contains 3 open reading frames, which		
CC	encode proteins AAY77177-v77178 and AAY77199. The vep ORF 1 protein is		
CC	defined in the specification as a PHA monomer synthase. (Updated on 15-		
CC	SEP-2003 to standardise OS field)		
XX			
SQ	Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;		
	Query Match	61.4%; Score 307.4; DB 3; Length 15872;	
	Best Local Similarity	76.9%; Pred. No. 1.7e-39;	
	Matches 390; Conservative	0; Mismatches 111; Indels 6; Gaps 1;	
Qy	1 TCGCTTCGTGGTGCAGTGACATCGCGGTGCAGGCCCTGCAGCGCGCGAGTGTCC	60	
Dd	9017 TCCTCCTCGTGGTGGCGCTGCATCTGCGGTACGGGCGTTCGGCACGCGAGTGGGG	9076	
Qy	61 ATGCGCGTTCGGCGCGCGGTGCATGCTGATGGGCACGGTCGAGACGTTTCGTCAGTTCTCG	120	
Dd	9077 CTCGCGCTGGCGCGGGGTGGCGGTGATGGCCCATCGGCGGGTTCGTGAGTTCTCC	9136	
Qy	121 CGGACGCGCGGCTGGCCCCCAACGGCGGTGCAGCGGTTCGCCGACGGCGGACGCGC	180	

Db	9137	CGGCAGAGAGGGCTGGCCGCGGACGCGCGTTCGAGAGCGTTCCTGGCGCGCGCGGACGGC	9139
QY	181	ACCGGGCTGGTCCGAGAGCGCTCGGGCTGCTCTGGTGGAGCGGCTGTCCGAGGCTCAAGCGT	240
Db	9197	ACCGGGCTGGGCGAGAGCGCTCGGCGTCTCTGGAGCGGCTGTTCGGACGCGCGCGC	9256
QY	241	CGCGGGCACAGAGTCTTCGCGCTGGTTCGCGGGTTCGCGCGTCAACTCCGACGCGCGCTCG	300
Db	9257	GCGGGGCACACGGTCTCTCGGCTTGGTTCACCGGCACCGCGCTCAACAGGACGGTGCCTCC	9316
QY	301	AACGGCTTGACGGCCCGGACGCGCGTCCAGCAGCGGTGATCCGCAAGGCACCTGGCC	360
Db	9317	AACGGGCTGACCGCGCCCAACGGCCGAGCCAGCAAGCGTCAATCGCCGAGGCGCTCGCC	9376
QY	361	GCGCCCGGACTGTCCACATCGAGCGTCCGACGCGGTGGAGCGCACCGGACCGGACGACC	420
Db	9377	GACGCCGGCTGTCCCGGGAGAGCTGGACGCGGTTCGAGGCGCACCGCACCGCGC	9436
QY	421	CTGGGGACCCGATCGAGGCGCGAGGCGCTCTGCGGCACCTACGGCCAGAAC	474
Db	9437	CTGGGCGACCCATCGAGGCGCGGGCGTCTCGCGCGCTTCGCGACGGAACGTTCCGGC	9496
QY	475	GAAACGCGCTGGGCTCGGCTCGGTG	501
Db	9497	GACACCCGCTGTGGCTTCGGCTCGCTG	9523
RESULT 13			
AAT80413			
ID	AAT80413 standard; DNA; 43280 BP.		
XX	AAT80413;		
DT	27-FEB-1998 (first entry)		
DE	Ty lactone synthase gene cluster.		
XX	Ty lactone synthase gene cluster; tylG gene; multifunctional protein;		
KW	polyketide; ty lactone synthesis; antibiotic; tylosin; ss.		
XX	Streptomyces fradiae.		
OS	Streptomyces fradiae.		
XX	Location/Qualifiers		
FH	Key	816..14243	
FT	CDS	/*tag= a	
FT	FT	/transl_except= (pos: 816..818, aa: Met)	
FT	FT	/note= "ORF1 encodes protein shown in AAW22601"	
FT	CDS	14351..13945	
FT	FT	/*tag= b	
FT	FT	/transl_except= (pos: 14351..14353, aa: Met)	
FT	FT	/note= "ORF2 encodes protein shown in AAW22602"	
FT	CDS	20010..31199	
FT	FT	/*tag= c	
FT	FT	/transl_except= (pos: 20010..20012, aa: Met)	
FT	FT	/note= "ORF3 encodes protein shown in AAW22603"	
FT	CDS	31232..36067	
FT	FT	/*tag= d	
FT	FT	/note= "ORF4 encodes protein shown in AAW22604"	
FT	CDS	36249..41774	
FT	FT	/*tag= e	
FT	FT	/note= "ORF5 encodes protein shown in AAW22605"	
XX	EP791655-A2.		
PN	XX		
PD	27-AUG-1997.		
XX	XX		
PF	19-FEB-1997;	97EP-00301056.	
XX	XX		
PR	22-FEB-1996;	96US-0012078P.	
XX	XX		
PA	(ELIL ) LILLY & CO ELI.		
XX	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;		





Db 17934 CTGGCCCTCGCGCGCGCCACCGCTCTTGGCCACGCGCGGCATGTTTACCGAGTTCTCG 17993  
QY 121 CGGACGCGGGGTGCGCCCGGACGCGCGCTGCAAGGCGTTGCCGACGCGCGGACGGC 180  
Db 17994 CGGACGCGGGGTGCGCCCGGACGCGCGCTGCAAGGCGTTGCCGACGCGCGGACGGC 18053  
QY 181 ACCGGCTGTGTCGAGGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 240  
Db 18054 ACCGGCTGTGTCGAGGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 18113  
QY 241 CGCGGACACGAGTCTCTGCGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 300  
Db 18114 AAGGGGACGCGCGCTCTCTGCGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 18173  
QY 301 AAGGGGACGCGCGCTCTCTGCGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 360  
Db 18174 AAGGGGACGCGCGCTCTCTGCGCGGTGCGCGCTGCTCTGTTGAGCGGCTGTCGAGGCTCAAGCGT 18233  
QY 361 GCGCGCGGACTGTCCATCGGACGCTCGACGCGGTGAGCGCGCACGCGCACCGGACGAC 420  
Db 18234 GCGCGCGGACTGTCCATCGGACGCTCGACGCGGTGAGCGCGCACGCGCACCGGACGAC 18293  
QY 421 CTGGCGACCGGATCGAGGCGCGGCGGCTGCTGCGCGGCTGCTGCGCGGCTGCTGCGCGG 474  
Db 18294 CTGGCGACCGGATCGAGGCGCGGCGGCTGCTGCGCGGCTGCTGCGCGGCTGCTGCGCGG 18353  
QY 475 GAACGCGCGGTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 501  
Db 18354 GAGCGCGGTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 18380

## RESULT 15

AAT68715

ID AAT68715 standard; DNA; 15872 BP.

XX AC

XX AAT68715;

XX DT

XX 01-SEP-1997 (first entry)

XX DE

XX Streptomyces venezuelae polyketide synthase vep ORF1.

XX KW

XX Polyketide synthase; polyhydroxyalkanoate monomer synthase;

XX KM

XX polyhydroxybutyrate; biodegradable polymer; vep gene;

XX KW

XX metabolic engineering; ss.

XX OS

XX Streptomyces venezuelae.

XX FH

XX Key Location/Qualifiers

XX CDS

XX 20..13912

XX /\*tag= a

XX CDS

XX 14056..14136

XX /\*tag= b

XX CDS

XX 14148..15827

XX /\*tag= c

XX FT

XX FT

XX FT

XX FT

XX PN

XX W09722711-A1.

XX XX

XX 26-JUN-1997.

XX XX

XX 18-DEC-1996; 96WO-US020119.

XX PF

XX 19-DEC-1995; 95US-0008847P.

XX XX

XX (MINU ) UNIV MINNESOTA.

XX FA

XX Sherman DH, Williams MD, Xue Y;

XX FI

XX WPI; 1997-341701/31.

XX DR

XX P-PSDB; AAW19629, AAW19630, AAW00918.

XX DR

XX Expression cassettes for production of poly:hydroxy:alkanoate(s) -

XX FT

XX provide wide range of biodegradable polymers for medical or industrial

XX use.

XX PT

XX PS  
XX Claim 54; Fig 23; 91pp; English.

XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide  
XX synthase (PKS) gene cluster encoding a polyene of 12 carbons (see also  
XX AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5' loading  
XX module and a 3' end domain. Each of the sequenced modules includes a keto  
XX -ACP, an acyltransferase, a dehydratase, a keto- reductase and an acyl  
XX carrier protein domain. The gene cluster was cloned using a heterologous  
XX hybridisation strategy from a genomic DNA library. A novel expression  
XX cassette encoding the first module from the vep gene cluster and module 7  
XX from the Streptomyces tyip gene cluster has polyhydroxyalkanoate (PHA)  
XX monomer synthase activity and can be used for PHA prodn. in host (esp.  
XX insect) cells for use as a biodegradable polymer

XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 U; 0 Other;  
XX

Query Match 61.0%; Score 305.8; DB 2; Length 15872;

Best Local Similarity 76.7%; Pred. No. 2.9e-39;

Matches 389; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 1 TCGTCTTGGTGTGCGGCTGCACATGGCGGTGCGGCTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 60  
Db 9017 TCGTCTTGGTGTGCGGCTGCACATGGCGGTGCGGCTGCAGGCCCTGCAGCGCGCGAGTGTCTCC 9076

QY 61 ATGCGCGCTCGCGGCGCGGCTGCATGTTGATGGGCGACGCGTTCGAGACGTTTCGTTCGAGTTTCG 120  
Db 9077 CTGCGGCTGCGGCGCGGCTGCATGTTGATGGGCGACGCGTTCGAGACGTTTCGTTCGAGTTTCG 9136

QY 121 CGGACGCGCGGCTGCGGCGCGGCTGCAGCGCGCTGCAGCGGCTTCGCGACGCGCGGCGGCGC 180  
Db 9137 CGGACGCGCGGCTGCGGCGCGGCTGCAGCGCGCTTCGCGACGCGCGGCGGCGC 9196

QY 181 ACCGCGTGTGCGGCGCGGCTGCGGCGCTGCTCTGTTGAGCGGCTTCGAGGCTTCAGGT 240  
Db 9197 ACCGCGTGTGCGGCGCGGCTGCGGCGCTGCTCTGTTGAGCGGCTTCGAGGCTTCAGGT 9256

QY 241 CGCGGCGCACGAGTCTTCGCGCGGCTGCGGCGGCTGCGGCGGCTCAACTCCGACGCGCGCTCG 300  
Db 9257 GCGGCGCACGAGTCTTCGCGCGGCTGCGGCGGCTGCGGCGGCTCAACTCCGACGCGCGCTCG 9316

QY 301 AAGCGCTTGACGCGCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 360  
Db 9317 AAGCGCTTGACGCGCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 9376

QY 361 GCGCGCGGCTGCTCCACATCGGACGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTG 420  
Db 9377 GACCGCGGCTGCTCCCGGAGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 9436

QY 421 CTGGCGGACCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 474  
Db 9437 CTGGCGGACCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 9496

QY 475 GAACGCGCGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 501  
Db 9497 GACCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 9523

Search completed: June 16, 2004, 23:46:41

Job time : 289 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:22:28 ; Search time 54.6667 seconds  
(without alignments)  
5085.922 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tcgttcgtcggtcgcgct.....csgtgcgtcggtcggtcggtg 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	314.4	62.8	77536	4	US-09-410-551B-1
2	307.4	61.4	15872	3	US-09-105-537-1
3	307.4	61.4	15872	4	US-09-091-609-1
4	307.4	61.4	15872	4	US-09-091-609-3
5	307.4	61.4	43280	2	US-08-804-227C-1
6	305.8	61.0	38506	3	US-09-320-878-19
7	305.8	61.0	38506	4	US-09-141-908-1
8	305.8	61.0	38506	4	US-09-657-440-19
9	304.2	60.7	4674	4	US-09-410-551B-26
10	304.2	60.7	4725	4	US-09-410-551B-24
11	304.2	60.7	4737	4	US-09-410-551B-30
12	304.2	60.7	4767	4	US-09-410-551B-28
13	304.2	60.7	4818	4	US-09-410-551B-32
14	304.2	60.7	13842	3	US-09-105-537-30
15	304.2	60.7	36778	3	US-09-105-537-5
16	303.2	60.5	77536	4	US-09-410-551B-1
17	294.2	58.7	1419	4	US-09-434-288-2
18	293	58.5	4466	4	US-09-410-551B-20
19	293	58.5	4478	4	US-09-410-551B-16
20	293	58.5	4547	4	US-09-410-551B-22
21	293	58.5	4571	4	US-09-410-551B-18
22	292.8	58.4	44377	2	US-08-804-227C-7
23	292.8	58.4	44377	2	US-08-804-198-1
24	288.8	57.6	4403765	3	US-09-103-840A-2
25	288.8	57.6	4411529	3	US-09-103-840A-1
26	288.2	57.5	20235	1	US-07-642-734C-3
27	288.2	57.5	20235	3	US-08-439-009A-3

28	285	56.9	11220	3	US-09-105-537-32	Sequence 32, Appl
29	283.2	56.5	11987	2	US-08-804-227C-13	Sequence 13, Appl
30	280.8	56.0	11219	1	US-07-642-734C-1	Sequence 1, Appl
31	280.8	56.0	11219	3	US-08-439-009A-1	Sequence 1, Appl
32	280.2	55.9	3978	4	US-09-266-965-19	Sequence 19, Appl
33	280.2	55.9	12249	4	US-09-266-965-74	Sequence 74, Appl
34	280.2	55.9	18331	4	US-09-266-965-96	Sequence 96, Appl
35	278.8	55.6	80161	3	US-09-036-987A-1	Sequence 1, Appl
36	278.8	55.6	80161	3	US-09-370-700-1	Sequence 1, Appl
37	278.8	55.6	80161	4	US-09-603-207-1	Sequence 1, Appl
38	277	55.3	4689	3	US-09-105-537-34	Sequence 34, Appl
39	273.8	54.7	50937	3	US-09-428-517-1	Sequence 1, Appl
40	272.2	54.3	47981	4	US-09-679-279-1	Sequence 1, Appl
41	267.4	53.4	4041	3	US-09-105-537-36	Sequence 36, Appl
C 42	249.4	49.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl
43	249	49.7	33529	3	US-09-144-085-3	Sequence 3, Appl
44	246.6	49.2	49377	1	US-08-764-233A-1	Sequence 1, Appl
C 45	246.2	49.1	4403765	3	US-09-103-840A-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-410-551B-1/c  
; Sequence 1, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1998-10-02  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 77536  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (52275)...(71465)  
US-09-410-551B-1

Query Match	62.8%	Score	314.4	DB	4	Length	77536
Best Local Similarity	76.8%	Pred. No.	1.1e-48	Indels	0	Gaps	0
Matches	384	Conservative	0	Mismatches	116		
QY	1	TCGTCCTCGCTGCGTGCACATGCGCGGTGAGGCGCTTGCAGCGCGCGGAGTGCTCC	60				
Db	42592	TCGTCCTCGCTGCGTGCACATGCGCGGTGAGGCGCTTGCAGCGCGCGGAGTGCTCC	42533				
QY	61	ATGCGCGTCCGCGCGCGCGTGTGATGGCAGCGTTCGACAGCTTCGTCGATTCCTCG	120				
Db	42532	CTGGCGTGTGCGCGGTGTCACCGTTCATGCGCAACACCGCAGAGCTCGTGGAGTTCTCC	42473				
QY	121	CGGACGCGCGGCTGGCCCCCGGCGCGTTCGCAAGCGTTCGCGCGCGGCGGAGCGC	180				
Db	42472	CGGACGCGCGGACTCTCCCGGACGCGCGTTCGCGCGCGGCGGAGCGG	42413				
QY	181	ACCGGCTGTGTCGAGGCGCGTCTCTCTGTGTGAGCGGCTGTCCGAGGCTCAGCGT	240				

Db 42412 ACCGCTTCGCGAGGCGTGGCGTTCTCGTGTGGAGGGCTCTCCGAGCGCGGCGC 42353  
QY 241 CCGCGGACACAGGTCTCTGCGGTGTGTCGCGGGTTCGCGGTCAACTCCGACGCGCGGTG 300  
Db 42352 AACGGGACACCGTTCTGCGGTGTGTCGCGGTTCGCGGTGAACACGAGACGTGCTCC 42293  
QY 301 AACGGCTTGACGGCCCGGACCGCGCTCCAGCAGCGGTGATCCGAAAGCACTGGCC 360  
Db 42292 AACGGAATCTCCGCCCGCCAGCGGTGCGCCAGCAGCGGTGATCCGAGCGGTGTGTC 42233  
QY 361 GCGCGCGGACTCTCCACATCGGACGTGACGCGGTGGAGCGCACGCGACCGGACGACC 420  
Db 42232 AACGCGGACTGCGCGCGCCGATGTGGACGTGGTGGAGCGCACGCGACCGCGCGG 42173  
QY 421 CTGGGGGACCGATAGCGCCGAGGGCTGTGGCCACCTAGCGGCAGAACCGGGAAGC 480  
Db 42172 CTGGGGGACCGATAGCGCGCAGGCGGTCTCTCGCGGCTACGCGGACGACCGCACG 42113  
QY 481 CCGCTGTGCTCGGTCGCT 500  
Db 42112 CCGCTACCTCGGTTGCT 42093

## RESULT 2

US-09-105-537-1  
; Sequence 1, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438051  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match 61.4%; Score 307.4; DB 3; Length 15872;  
Best Local Similarity 76.9%; Pred. No. 2e-47;  
Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCGGTGCGATGCGGTGCGCGCTGCGCGCTGCGCGCGGAGTGTCC 60  
Db 9017 TCTCTCTGCTGTGGCTGTGATCTGGCGGTACGGCGGTGCGGACGCGGAGTGGGG 9076  
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTGAAGGACCGGTGAGAGTTCGTCGAGTTCTCG 120  
Db 9077 CTGCGCTGCGCGCGCGGTGCGGTGATGGCGATCGCGCGGTTCGTGGAGTTCTCC 9136  
QY 121 CGGACGCGCGGTGCGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 180  
Db 9137 CGGACGCGCGGTGCGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9136  
QY 181 ACCGCTGTGTCGAGGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 240  
Db 9197 ACCGCTGTGTCGAGGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9256  
QY 241 CCGCGGACACAGGTCTCTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 300  
Db 9257 GCGGGGACACGCTCTCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9316  
QY 301 AACGGCTTGACGGCCCGGACCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 360  
Db 9317 AACGGCTTGACGGCCCGGACCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9376

QY 361 GCGCGCGGACTGTCCACATCGGACGTGCGGTGAGGCGCACGCGCACCGGAGCGACC 420  
Db 9377 GACGCGGCGGTGTCCCGGAGGACGTGCAOGGCTGAGGCGCACGCGCACCGGACCGCG 9436  
QY 421 CTGGGCGACCGATAGCGCGGAGCGGTGCTGGCCACTAGGCGCAAGAAC-----GG 474  
Db 9437 CTGGGCGACCGATAGCGCGGCGGTGCTGGCGCTCGGACGGAACCGTTCCGCGC 9496  
QY 475 GAAACGCGCTGTGCTCGGTCGCTG 501  
Db 9497 GACCAACCGCTGTGCTCGGTCGCTG 9523  
RESULT 3  
US-09-091-609-1  
; Sequence 1, Application US/09091609  
; Patent No. 6600029  
; GENERAL INFORMATION:  
; APPLICANT: SHERMAN, DAVID H.  
; APPLICANT: WILLIAMS, MARK D.  
; APPLICANT: XUE, YONGQUAN  
; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
; TITLE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES  
; FILE REFERENCE: 600.297052  
; CURRENT APPLICATION NUMBER: US/09/091,609  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: PCT/US96/20119  
; EARLIER FILING DATE: 1996-12-18  
; EARLIER APPLICATION NUMBER: 60/008,847  
; EARLIER FILING DATE: 1995-12-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (20)....(13909)  
US-09-091-609-1

Query Match 61.4%; Score 307.4; DB 4; Length 15872;  
Best Local Similarity 76.9%; Pred. No. 2e-47;  
Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCGGTGCGATGCGGTGCGCGCTGCGCGCTGCGCGCGGAGTGTCC 60  
Db 9017 TCTCTCTGCTGTGGCTGTGATCTGGCGGTACGGCGGTGCGGACGCGGAGTGGGG 9076  
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTGAAGGACCGGTGAGAGTTCGTCGAGTTCTCG 120  
Db 9077 CTGCGCTGCGCGCGCGGTGCGGTGATGGCGATCGCGCGGTTCGTGGAGTTCTCC 9136  
QY 121 CGGACGCGCGGTGCGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 180  
Db 9137 CGGACGCGCGGTGCGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9136  
QY 181 ACCGCTGTGTCGAGGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 240  
Db 9197 ACCGCTGTGTCGAGGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9256  
QY 241 CCGCGGACACAGGTCTCTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 300  
Db 9257 GCGGGGACACAGGTCTCTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9316  
QY 301 AACGGCTTGACGGCCCGGACCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 360  
Db 9317 AACGGCTTGACGGCCCGGACCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 9376  
QY 361 GCGCGCGGACTGTCCACATCGGACGTGCGGTGAGGCGCACGCGCACCGGAGCGACC 420  
Db 9377 GACGCGGCGGTGTCCCGGAGGACGTGGAGCGCACGCGCACCGGACCGGACCGCGC 9436



QY 121 CGGAGCGGGGCTGCCCCCGACGGCCGCTGCAAGGGCTTCGCGACGGCGGGACGGC 180  
Db |||||  
QY 36999 CGGCAAGGGGCTGGCCCGCGACGGCCGCTGCAAGGGCTTCGCGCCCGCGCGACGGC 37058  
Db |||||  
QY 181 ACCGGTGTTCGAGGGCGTCGGGCTGCTCTCTGGTGGAGCGGTGTCCAGGCTCAGCGT 240  
Db |||||  
QY 37059 ACCGGTGGCGCGAGGGCTGCGGCTGCTCTCTGGTGGAGCGGTGTCCAGGCTCAGCGT 37118  
Db |||||  
QY 241 CGGCGGACACAGTCTCTCGCTGTCTCGGGTCCGGGTGCGGCTCAACTCCAGCGCGGTCCG 300  
Db |||||  
QY 37119 GCGGGGACACAGTCTCTCGCTGTCTCGGGTCCGGGTGCGGCTCAACTCCAGCGCGGTCCG 37178  
Db |||||  
QY 301 AACGGCTTACGCGCCCGGACCGCGCTCCAGCGCGGTGTATCCGCAAGGCACTTGGCC 360  
Db |||||  
QY 37179 AACGGCTTACGCGCGCCGACCGCGCTCCAGCGCGGTGTATCCGCAAGGCACTTGGCC 37238  
Db |||||  
QY 361 GCGCGCGGACTGTCCACATCGACGCTGCGCGGTGGAGCGGCGACCGGACCGGACGACC 420  
Db |||||  
QY 37239 GACGCGGGCTGTCCCGCGAGGACGTGGAGCGGTGCGAGCGGCGACCGGACCGGACCGG 37298  
Db |||||  
QY 421 CTGGGGACCCGATCGAGCGCGAGGGCTGCTGGCCACTACGCGCGACGACC-----GG 474  
Db |||||  
QY 37299 CTGGGGACCCGATCGAGCGCGGGGGCTGCTGGCGCGCTCCGCGACGGAACGTTCCGCG 37358  
Db |||||  
QY 475 GAAACGCGCGCTGTGGCTCGGCTCGGTG 501  
Db |||||  
QY 37359 GACCGCGCGCTGTGGCTCGGCTCGGTG 37385  
Db |||||

## RESULT 6

US-09-320-878-19  
; Sequence 19, Application US/09320878A  
; Patent No. 6117659

; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A  
; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139  
; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080  
; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 19  
; LENGTH: 38506

; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae

US-09-320-878-19

Query Match 61.0%; Score 305.8; DB 3; Length 38506;  
Best Local Similarity 76.7%; Pred. No. 3.8e-47;

Matches 389; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCGGTGCACATGGCGGTGCAGGCGCTTCGAGCGCGCGAGTCTCC 60  
Db |||||

QY 3727 TCGTCTGCTGTGCTGCCCTGCACTCGCGGTGCAAGGCGCGAGGTCGAC 3786  
Db |||||

QY 61 ATGGCGCTCGCGGGCGGCTGATGGTGCAGCGGTGCGGACCGTTCGTGAGTTCG 120  
Db |||||

Db 3787 ATGGCGCTCGCGGGCGGCTGATGCCACGCGCGGATGTTCTCGAGTTCAGC 3846  
QY |||||  
QY 121 CGGCAAGGGGCTGGCCCGCGACGGCCGCTGCAAGGGCTTCGCGCCCGCGCGACGGC 180  
Db |||||  
QY 3847 CGGCAAGGGGCTGGCCCGCGACGGCCGCTGCAAGGGCTTCGCGCCCGCGCGACGGC 3906  
Db |||||  
QY 181 ACCGGTGTTCGAGGGCGTCGGGCTGCTCTCTGGTGGAGCGGTGTCCAGGCTCAGCGT 240  
Db |||||  
QY 3907 ACCAGTGTTCGAGGGCGTCGGGCTGCTCTCTGGTGGAGCGGTGTCCAGGCTCAGCGT 3966  
Db |||||  
QY 241 CGGCGGACACAGTCTCTCGCGCTGCTCGCGGTGCGGCTCAACTCCAGCGCGGTCCG 300  
Db |||||  
QY 3967 AACGGACACAGTCTCTCGCGCTGCTCGCGCGGCGGCTGAACAGGACGCGCGGACG 4026  
Db |||||  
QY 301 AACGGTTCAGCGCCCGGACCGCGCTCCAGCGCGGTGTATCCGCAAGGCACTTGGCC 360  
Db |||||  
QY 4027 AACGGCTTACGCGCGCCGACCGCGCTCCAGCGCGGTGTATCCGCAAGGCACTTGGCC 4086  
Db |||||  
QY 361 GCGCGCGGACTGTCCACATCGGACGCTGCGAGCGGTGGAGCGGCGACCGGACCGACC 420  
Db |||||  
QY 4087 GACGCGGGCTGACGACCTCCGACGCTGGAGCTGCTCGAGCGGCGACCGGACCGGCG 4146  
Db |||||  
QY 421 CTGGGGACCCGATCGAGCGCGAGGGCTGCTGGCCACTACGCGCGACGACC-----G 474  
Db |||||  
QY 4147 CTGGGGACCCGATCGAGCGCGGCGGCTGATCGCCACTACGCGCGGCGGTGCGACG 4206  
Db |||||  
QY 475 GAAACGCGCGCTGTGGCTCGGCTCGGTG 501  
Db |||||  
QY 4207 GAACAGCGCGCTGCGCTCGGCTCGGTG 4233  
Db |||||

## RESULT 7

US-09-141-908-1

; Sequence 1, Application US/09141908

; Patent No. 6503741

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
; FILE REFERENCE: 300622002100

; CURRENT APPLICATION NUMBER: US/09/141,908  
; CURRENT FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: PROV. 60/076,919  
; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: PROV. 60/087,080  
; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-141-908-1

Query Match 61.0%; Score 305.8; DB 4; Length 38506;  
Best Local Similarity 76.7%; Pred. No. 3.8e-47;

Matches 389; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCGGTGCACATGGCGGTGCAGGCGCTTCGAGCGCGCGAGTCTCC 60  
Db |||||

QY 3727 TCGTCTGCTGTGCTGCCCTGCACTCGCGGTGCAAGGCGCGAGGTCGAC 3786  
Db |||||

QY 61 ATGGCGCTCGCGGGCGGCTGATGGTGCAGCGGTGCGGACCGTTCGTGAGTTCG 120  
Db |||||

3787	ATGGCGTCCCGCGCGCGTGGCGGTGATGCCACGCGCGGATGTTCTGCGAGTTTCAGC	3846
QY		
121	CGCGAGCGCGGGCTGGCCCCCGACGCGCGCTGCAAGGCGTTTCGCGACGCGCGGACGGC	180
DB		
3847	CGCGAGCGCGGGCTGGCGGGGACGCGCGGTGCAAGGCGTTTCGCGCGGACGGC	3906
QY		
181	ACGGCTGTGTCAGAGGCGTCGGGCTCTCTCTGTGTGAGCGGCTGTCCGAGGCTCAGCGT	240
DB		
3907	ACGAGTGTGTCAGAGGCGTCGGCGTCTCTCTGTGTGAGCGCTGTCTGACGCGCCCGCGC	3966
QY		
241	CGCGGGCACCAAGTCTCTGCGCGTGTCCGCGGTTCGCGGTCAACTCCGACGCGCGTCTG	300
DB		
3967	AACGGACACCAAGTCTCTGCGGTGTCCGCGGCAGCGCGGTGAACAGGACGCGCGAGC	4026
QY		
301	AACGGCTTGACGCGCCCGAAGCGCGCGTCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC	360
DB		
4027	AACGGCTCTACGGCTCGAACAAGGCGCTTCGACAGCGCGCTCATCCGGCGCGCTGGCG	4086
QY		
361	GCGCGCGGACTGTCTCATCGAGCGTCTGACGCGGTGTGAGCGGACGACGACCGGGACGACC	420
DB		
4087	GACCGCGGTGACGACCTCTCGAGCTGGAAGTGTGTGAGGCAACGCGCACGGGCAACGCGA	4146
QY		
421	CTGGGCGACCCGATCGAGGCGGAGGCGTCTGTGCCACTTACGCGCCAGAACCG-----G	474
DB		
4147	CTCGGCGACCCGATCGAGGCGGAGGCGCTGATCGCCACCTTACGGCCAGGCGCGTGCAGC	4206
QY		
475	GAAACGCGCTGTGGTCTCGGTCGGTG	501
DB		
4207	GAAACGCGCTGTGGCTCTCGGTCGGTG	4233

```

RESULT 8
US-09-657-440-19
; Sequence 19, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-19

```

Qy	181	ACGGGTGTTCCGAGGCGTTCGGGCTGCTCTCTGGTGGAGCGGCGTGTCCGAGGCTCAGCGT	240
Db	3907	ACCAGTGTGTTCCGAGGGCGTTCGGCGTCTCTCTCGTCCGAGCGCTGTCCGACGCCCGCCGC	3966
Qy	241	CGCGGGCACCAGGTCTTCGCGTGTGTCGGGGGTTCGGCGGTCAACTCCGACCGCGCGTCG	300
Db	3967	AACGGCACACAGGTCTTCGCGTGTGTCGGCGGTCCGCGCGAGCGCGTGAACACGAGCAGCGCGGAGC	4026
Qy	301	AACGGCTTGACGGCCCCGGAACGGCCCCGTCCACAGCAGCGGTGATCCGCAAGCACTGSGCC	360
Db	4027	AACGGCTCTACGGCTCCGAAACGGGCCCTCCGACAGCGGTTCATCCGGCGCGCGCTGGCG	4086
Qy	361	GC CGCGCGACTGTCCACATCGAGCGTCCGACGGGTGGAGGCGCACGCGACCGGGACGACC	420
Db	4087	GACGCCCGGTGACGACCTCCGACGTGGAGCTGTCTGAGGGCACACGGCACCGGCACCGCA	4146
Qy	421	CTCGGGCGACCGATCCAGCGCCGAGGCGCTGCTGGCCACCTACGGCCAGAACCG-----G	474
Db	4147	CTCGGGCGACCGGATCCAGGCGCAGGCCGCTGATCGCCACCTACGGCCAGGGCCGTGACGAC	4206
Qy	475	GAACACCGCTGTGGCTTCGGTTCGGTG	501
Db	4207	GAACACCGCTGCGCCTTCGGTTCGGTG	4233

RESULT 9  
 US-09-410-551B-26  
 ; Sequence 26, Application US/09410551B  
 ; Patent No. 6503737

RESIST 9

```

US-09-410-551B-26
; Sequence 26, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 4674
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4673)
US-09-410-551B-26

```

Qy	1	TCGCTTTGCTGTGCTGCGCTGCACATGCGCGGTGAGGCCTTCAGCGCGCGGAGTGTCTCC	60
Db	3727	TCGTCGTGCTGTGCTGCGCTGCACCTTCGCGGTGCAGGCCCTTCGCGAAGGCGGAGGTGCAC	3786
Qy	61	ATGCGCTTCGCGGCGCGCTGATGGTGTATGGGCACGGTTCGAGAGTTCTGCTGAGTTCTCG	120
Db	3787	ATGCGCTTCGCGGCGCGGTGGCGGTGATGCCACGCGCGGATGTTCTGCTGAGTTTCAGC	3846
Qy	121	CGGCAAGCGCGGGCTGGCGCCCGAAGCGCGCTGCAGCGCGTTTCGCCGACGGCGCGGACGGC	180
Db	3847	CGGCAAGCGCGGCTGGCGCGGGAAGCGCGGTTCGAGAGCGTTTCGCCGCGTTCGCGCGACGGC	3906

QY 121 CGGAGCGCGGGCTGGCCCCCGACCGCGCGCTCGAAGGGGCTTCGCCGACGCGCGCGACGGC 180  
DB 1092 CGGAGCGCGGGCTCGCGCCGACCGCGCGCGCTCGAAGGGGCTTCGCCGCGCGCGCGACGGT 1151  
QY 181 ACCGGCTGGTCCGAGGGGCTGGGGCTGCTCTGCTGGTGGAGCGGCTCTCGAGGCTCAGGCT 240  
DB 1152 ACGAGCTTCGCGAGGGGCGCGGGTGGCTTGGCTGGAGCGGCTCTCGAGCGGAGCGC 1211  
QY 241 CGCGGGCACCGAGTCTCTCGCCCTGGTTCGCGGGTTCGCGGGTCAACTTCGACGCGCGGGTGG 300  
DB 1212 CACGGCCACACCGCTCTCGCCCTGGTTCGCGGGTTCGCGGGTCAACTTCGCGAGCGCGGCTGG 1271  
QY 301 AAGCGCTTACCGCGCCCGACCGCGCGCTCCGAGCGCGGTCGAGCGCGGTCGAGCGGCTCGG 360  
DB 1272 AAGCGTCTGTGCGCGCGGACCGCGCGCTCCGAGCGCGGTCGAGCGCGGTCGAGCGGCTCGG 1331  
QY 361 GCGCGCGGACTCTCCACATCGACGCTCGACGCGGTCGAGCGCGGTCGAGCGCGGTCGAGCGG 420  
DB 1332 AAGCGGAACTCACCCTCGCGCGGTCGAGCGGTCGAGCGGTCGAGCGGTCGAGCGGTCGAG 1391  
QY 421 CTGGGCGACCGCGATCGAGCGCGGCGCTGCTGGCGACCTACGCGCGAGAACCGGAAACG 480  
DB 1392 CTCGGCGACCGCATCGAGCGCGGCGGCTGCTCGGACGCTACGCGAGCGGTCGAGCGGCG 1451  
QY 481 CGCGTGTGGCTCGGGTTCGGT 501  
DB 1452 CCCCTGCTGCTCGGGTTCGGT 1472

RESULT 10  
US-09-410-551B-24  
; Sequence 24, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 4725  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(4724)  
US-09-410-551B-24

Query Match 60.7%; Score 304.2; DB 4; Length 4725;  
Best Local Similarity 75.4%; Pred. No. 7.8e-47;  
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 1 TCGTCTTCGCTGGCTGCGCTGCGATGGCGGTGCGAGCGCGCGCGAGTGTCTCC 60  
DB 972 TCGTCTTCGCTGGCTGCGCTGCGATGGCGGTGCGAGCGCGCGAGTGTCTCC 1031

QY 61 ATGGCGCTCGCGGGCGCGGTGATGCTGATGGGCAAGGTCGAGACGCTCGTTCGAGTCTCCG 120  
DB 1032 CTCGCCCTGGTTCGGCGGTGTACCGGTGATGGCGTCCGCCGCGGAGTTCGTCGAGTCTCC 1091  
QY 121 CGGAGCGCGGGCTGGCCCCCGACCGCGCGCTCGAAGGGGCTTCGCCGACGCGCGCGACGGC 180  
DB 1092 CGGAGCGCGGGCTGGCCCCCGACCGCGCGCTCGAAGGGGCTTCGCCGCGCGCGAGCGT 1151  
QY 181 ACCGGCTGGTTCGAGGGGCTCGGGCTGCTCTGCTGGTGGAGCGGCTCTCGAGGCTCAGGCT 240  
DB 1152 ACGAGCTTCGCGAGGGGCGCGGGTGGCTTGGCTGGAGCGGCTCTCGAGCGGAGCGC 1211  
QY 241 CGCGGGCACCGAGTCTCTCGCCCTGGTTCGCGGGTTCGCGGGTCAACTTCGACGCGCGGGTGG 300  
DB 1212 CACGGCCACACCGCTCTCGCCCTGGTTCGCGGGTTCGCGGGTCAACTTCGCGAGCGCGGCTGG 1271  
QY 301 AAGCGCTTACCGCGCCCGACCGCGCGCTCCGAGCGCGGTCGAGCGCGGTCGAGCGGCTCGG 360  
DB 1272 AAGCGTCTGTGCGCGCGGACCGCGCGCTCCGAGCGCGGTCGAGCGGTCGAGCGGCTCGG 1331  
QY 361 GCGCGCGGACTCTCCACATCGACGCTCGACGCGGTCGAGCGCGGTCGAGCGCGGTCGAGCGG 420  
DB 1332 AAGCGGAACTCACCCTCGCGCGGTCGAGCGGTCGAGCGGTCGAGCGGTCGAGCGGTCGAG 1391  
QY 421 CTGGGCGACCGCGATCGAGCGCGGCGCTGCTGGCGACCTACGCGCGAGAACCGGAAACG 480  
DB 1392 CTCGGCGACCGCATCGAGCGCGGCGGCTGCTCGGACGCTACGCGAGCGGTCGAGCGGCG 1451  
QY 481 CGCGTGTGGCTCGGGTTCGGT 501  
DB 1452 CCCCTGCTGCTCGGGTTCGGT 1472

RESULT 11  
US-09-410-551B-30  
; Sequence 30, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 4737  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(4736)  
US-09-410-551B-30

Query Match 60.7%; Score 304.2; DB 4; Length 4737;  
Best Local Similarity 75.4%; Pred. No. 7.8e-47;  
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
QY 1 TCGTCTTGGTGGTGGCTCCATGCGCGTGCAGGCGCTTGCAGCGCGGCGAGTGTCC 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
972 TCGTCTTGGTGGTGGCTCCATGCGCGTGCAGGCGAGTGTCCATGCGTGGCGAGTGTCCG 1031
QY 61 ATGGCGCTCGCGGCGGCGTGTGATGGTATGGCAGGCTGCAGAGCTTGTGAGTGTCCG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1032 CTGCGCTTGGTGGGCGGCTGATGCGGTGATGGCTGCGCGCGGAGTGTGAGTGTCTCC 1091
QY 121 CGGAGCGCGGCGCTGGCGCCCGCGAGCGGCGCTGCAAGGCGTTTCGCGCGCGGCGAGCGC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1092 CGGAGCGCGGCGCTGGCGCCCGCGAGCGGCGCGGCGAGGCGTTTCGCGCGCGGCGAGCGT 1151
QY 181 ACCGCTTGGTCCGAGGCGCTGGCGCTGCTCTGTGTGAGAGCGCTTTCGAGGCTCAGGT 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1152 ACAGCTTTCGCGGCGGCGGCTGCGCTGCTGTGTGTGAGAGCGCTTTCGAGAGCGCTC 1211
QY 241 CGCGGCGACCAAGTCTTCCGCGTGTGCGCGGCTGGCGGCTCAACTCCGAGCGGCGCTCG 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1212 CACGCGCACACCGCTCTCGCCCTGTACGCGGCTCCGCGGCTAACTCCGAGCGGCGCTCG 1271
QY 301 AACGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1272 AACGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1331
QY 361 GCGCGGCGACTGTCCATATCGAGCGTGCAGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGG 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1332 AACGCGAACTCAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1391
QY 421 CTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1392 CTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 481 CCGCTGTGGCTCGGCTCGGCTG 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1452 CCGCTGTGGCTCGGCTCGGCTG 1472
```

## RESULT 12

```
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
```

## GENERAL INFORMATION:

```
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
```

## TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

## FILE REFERENCE: 30062-20026.00

## CURRENT APPLICATION NUMBER: US/09/410,551B

## CURRENT FILING DATE: 1999-10-01

## PRIOR APPLICATION NUMBER: US 60/139,650

## PRIOR FILING DATE: 1999-06-17

## PRIOR APPLICATION NUMBER: US 60/123,810

## PRIOR FILING DATE: 1999-03-11

## PRIOR APPLICATION NUMBER: US 60/102,748

## PRIOR FILING DATE: 1998-10-02

## NUMBER OF SEQ ID NOS: 72

## SOFTWARE: FastSeq for Windows Version 4.0

## SEQ ID NO 28

## LENGTH: 4767

## TYPE: DNA

## ORGANISM: Artificial Sequence

## FEATURE:

```
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
```

```
; OTHER INFORMATION: PKS synthase fragment
```

## FEATURE:

```
; NAME/KEY: CDS
```

```
; LOCATION: (3)..(4766)
```

```
US-09-410-551B-28
```

```
Query Match 60.7%; Score 304.2; DB 4; Length 4767;
Best Local Similarity 75.4%; Pred. No. 7.8e-47;
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1 TCGTCTTGGTGGTGGCTCCATGCGCGTGCAGGCGCTTGCAGCGCGGCGAGTGTCC 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
972 TCGTCTTGGTGGTGGCTCCATGCGCGTGCAGGCGAGTGTCCATGCGTGGCGAGTGTCCG 1031
QY 61 ATGGCGCTCGCGGCGGCGTGTGATGGTATGGCAGGCTGCAGAGCTTGTGAGTGTCCG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1032 CTGCGCTTGGTGGGCGGCTGATGCGGTGATGGCTGCGCGCGGAGTGTGAGTGTCTCC 1091
QY 121 CGGAGCGCGGCGCTGGCGCCCGCGAGCGGCGCTGCAAGGCGTTTCGCGCGCGGCGAGCGC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1092 CGGAGCGCGGCGCTGGCGCCCGCGAGCGGCGCGGCGAGGCGTTTCGCGCGCGGCGAGCGT 1151
QY 181 ACCGCTTGGTCCGAGGCGCTGGCGCTGCTCTGTGTGAGAGCGCTTTCGAGGCTCAGGT 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1152 ACAGCTTTCGCGGCGGCGGCTGCGCTGCTGTGTGTGAGAGCGCTTTCGAGAGCGCTC 1211
QY 241 CGCGGCGACCAAGTCTTCCGCGTGTGCGCGGCTGGCGGCTCAACTCCGAGCGGCGCTCG 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1212 CACGCGCACACCGCTCTCGCCCTGTACGCGGCTCCGCGGCTAACTCCGAGCGGCGCTCG 1271
QY 301 AACGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1272 AACGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1331
QY 361 GCGCGGCGACTGTCCATATCGAGCGTGCAGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGG 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1332 AACGCGAACTCAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1391
QY 421 CTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1392 CTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 481 CCGCTGTGGCTCGGCTCGGCTG 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1452 CCGCTGTGGCTCGGCTCGGCTG 1472
```

## RESULT 13

```
US-09-410-551B-32
```

```
; Sequence 32, Application US/09410551B
```

```
; Patent No. 6503737
```

## GENERAL INFORMATION:

```
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
```

## TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

## FILE REFERENCE: 30062-20026.00

## CURRENT APPLICATION NUMBER: US/09/410,551B

## CURRENT FILING DATE: 1999-10-01

## PRIOR APPLICATION NUMBER: US 60/139,650

## PRIOR FILING DATE: 1999-06-17

## PRIOR APPLICATION NUMBER: US 60/123,810

## PRIOR FILING DATE: 1999-03-11

## PRIOR APPLICATION NUMBER: US 60/102,748

## PRIOR FILING DATE: 1998-10-02

## NUMBER OF SEQ ID NOS: 72

## SOFTWARE: FastSeq for Windows Version 4.0

## SEQ ID NO 32

## LENGTH: 4818

## TYPE: DNA

## ORGANISM: Artificial Sequence

## FEATURE:

```
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
```

```
; OTHER INFORMATION: PKS synthase fragment
```

## FEATURE:



```
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-410-551B-32

Query Match      60.7%; Score 304.2; DB 4; Length 4818;
Best Local Similarity 75.4%; Pred. No. 7.8e-47;
Matches 378; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGCTGCGCTGCATCGCGGTGCAGCGCTGCAGCGCTGCAGCGCGAGTCTCC 60
Db 972 TCGTCTTCATGTCGCTTGCCTTGCACAGCGAGGAGTCTCCCTGCGCTCGGCGCGAATGCTCG 1031
QY 61 ATGCGCGCTCCGCGCGCGCTGATGGTGCATGGGCAACGCTCGAGAGTTCGTCGAGTTCCTCG 120
Db 1032 CTGCGCTGCTGCGCGCTGTCACGGTGCATGCGCTGCGCGCGGATTCGTCGAGTTCCTCC 1091
QY 121 CGGAGCGCGCGCTGGCGCGCGCGCTGCAAGCGCTGCGCGCGTTCGCGCGCGCGCGCGCG 180
Db 1092 CGGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1151
QY 181 ACCGGCTGCTCCAGCGCGCTGCGCTGCTCTCTGTTGAGCGGCTGCTCCGAGGCTCAGCGT 240
Db 1152 ACGAGCTTCGCGCGCGCGCGCTGCGCTGCTCTCTGTTGAGCGGCTGCTCCGAGGCTCAGCGT 1211
QY 241 CGCGCGCAACAGTCTCGCGCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 300
Db 1212 CAGCGCGCAACAGTCTCGCGCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1271
QY 301 AAGCGCTTGAGCGCGCGCGCGCGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 360
Db 1272 AAGCGCTTGAGCGCGCGCGCGCGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1331
QY 361 GCGCGCGAGTCTCCATCGGAGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 420
Db 1332 AAGCGCGAGTCTCCATCGGAGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1391
QY 421 CTGGCGCGAGCGCGCGCGCGCGCGCGCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 480
Db 1392 CTGGCGCGAGCGCGCGCGCGCGCGCGCTGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1451
QY 481 CCGCTGCTGCTGCGCGCTGCGCGT 501
Db 1452 CCGCTGCTGCTGCGCGCTGCGCGT 1472

RESULT 14
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match      60.7%; Score 304.2; DB 3; Length 13842;
Best Local Similarity 76.5%; Pred. No. 7.6e-47;
Matches 388; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCTGCGCTGCATCGCGGTGCAGCGCTGCAGCGCTGCAGCGCGAGTCTCC 60
Db 3844 TCGTCTTCGCTGCTGCGCTGCATCGCGGTGCAGCGCTGCAGCGCTGCAGCGCGAGTCTCC 3903
```

```
QY 61 ATGCGCGTTCGCGCGCGCTGATGGTATGGGCAACGCTCGAGACGTTTCGTCGAGTTCCTCG 120
Db 3904 ATGCGCGTTCGCGCGCGCTGATGGTATGGGCAACGCTCGAGTTCCTCGTCGAGTTCAGC 3963
QY 121 CGGAGCGCGCGCTGCGCGCGCGCTGCAAGCGCTTCGCGCGCGCGCGCGCGCGCGCG 180
Db 3964 CGGAGCGCGCGCTGCGCGCGCGCTGCAAGCGCTTCGCGCGCGCGCGCGCGCGCGCG 4023
QY 181 ACCGGCTGCTCCAGCGCGCTGCGCTGCTCTCTGTTGAGCGGCTGCTCCGAGGCTCAGCGT 240
Db 4024 ACCAGTGGTTCGAGCGCGCTGCGCTGCTCTCTCTGTTGAGCGGCTGCTCCGAGGCTCAG 4083
QY 241 CGGCGCGCGCGCTGCTGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCG 300
Db 4084 AACGGGACACAGGCTCTCGCGCTGCTGCGCGCGCGCGCGCTGCGCGCGCGCGCGCG 4143
QY 301 AAGCGCTTGAGCGCGCGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCG 360
Db 4144 AAGCGCTTGAGCGCGCGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCG 4203
QY 361 GCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCG 420
Db 4204 GAGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCG 4263
QY 421 CTGGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCG 474
Db 4264 CTGGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCG 4323
QY 475 GAAACGCGCGCTGCGCGCTGCGCGCTGCGCGT 501
Db 4324 GAAACGCGCGCTGCGCGCTGCGCGCTGCGCGT 4350

RESULT 15
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match      60.7%; Score 304.2; DB 3; Length 36778;
Best Local Similarity 76.5%; Pred. No. 7.3e-47;
Matches 388; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGCTGCGCTGCATCGCGGTGCAGCGCTGCAGCGCGCGAGTCTCC 60
Db 5585 TCGTCTTCGCTGCTGCGCTGCATCGCGGTGCAGCGCTGCAGCGCGCGAGTCTCC 5644
QY 61 ATGCGCGCTCGCGCGCGCTGATGGTATGGGCAACGCTCGAGACGTTTCGTCGAGTTCCTCG 120
Db 5645 ATGCGCGCTCGCGCGCGCTGATGGTATGGGCAACGCTCGAGTTCCTCGTCGAGTTCAGC 5704
QY 121 CGGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCG 180
Db 5705 CGGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCG 5764
QY 181 ACCGGCTGCTCCAGCGCGCTGCGCGCTGCTCTCTGTTGAGCGGCTGCTCCGAGGCTCAGCGT 240
```

```
Db 5765 ACCAGCTGGTCCGAGGGGCTCGGCGTCTCTCTCGTGGAGCGCTGTGGACGCCCGCGC 5824
QY 241 CGCGGGCACCGGTCTCTGCGGTGTTCTCGCGGTGCGGCTCACTCCGACGGCGCGTGG 300
Db 5825 AACGGACACCGAGTCTCTGCGGTCTCTGCGGTGCGGCTTGAACAGGACGGCGCGAGC 5884
QY 301 AACGGCTTGACGGCCCCGAAAGGCCCGTCCAGAGCGGTGATCCGCAAGGCACCTGGCC 360
Db 5885 AACGGCTTACGGCTCCGACGGGCCCTCGACAGCGCGTCTATCCGGCGCGCTGGCG 5944
QY 361 GCCGCCGGACTGTCCACATCGAACGTGACCGCGTGGAGGCGCACGGCACCGGGACGACC 420
Db 5945 GACGCCCGGCTGACACACCTCCGACGTGACGTCTGAGGACACGGCACGGGCACGCGA 6004
QY 421 CTGGGCGACCCGATCGAGGCCGAGGCGGTCTGTGGCCACCTACGGCCAGAACCG-----G 474
Db 6005 CTCGGCGACCCGATCGAGGCGCAGGCCCTGTGATCGCCACCTACGGCCAGGGCGCGTACGAC 6064
QY 475 GAAAGCCGCTGTGGCTCGGGTCGGTG 501
Db 6065 GACAGCCGCTGCGCCTCGGGTCGTTG 6091
```

Search completed: June 17, 2004, 03:21:55  
Job time : 65.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds  
(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501  
Sequence: 1 tcgtcttcgctggtcgct.....cgctggtcggtcggtcggtg 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications NA.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
  - 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
  - 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
  - 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.2	65.5	18438	15	US-10-156-761-2886
2	328.2	65.5	125746	15	US-10-156-761-15102
3	328.2	65.5	9025608	15	US-10-156-761-1
4	317	63.3	18435	15	US-10-156-761-412
5	317	63.3	100000	15	US-10-156-761-15103
6	316.6	63.2	23238	15	US-10-156-761-415
7	315.4	62.8	77536	10	US-09-940-316B-1
8	315.4	62.8	10056	15	US-10-156-761-411
9	310.2	61.9	10056	15	US-10-156-761-411
10	309.6	61.8	10692	15	US-10-156-761-414
11	309	61.7	10839	15	US-10-156-761-2882
12	309	61.7	86941	17	US-10-461-194-2
13	307.4	61.4	15872	9	US-09-861-289-1
14	307.4	61.4	15872	9	US-09-860-846-1

15	307.4	61.4	15872	10	US-09-988-384B-1	Sequence 1, Appli
16	307.4	61.4	15872	10	US-09-836-821-1	Sequence 1, Appli
17	307.4	61.4	15872	15	US-10-271-889-44	Sequence 44, Appl
18	305.8	61.0	38506	10	US-09-793-708-19	Sequence 19, Appl
19	305.8	61.0	38506	15	US-10-201-365-1	Sequence 1, Appli
20	305.8	61.0	38506	15	US-10-160-539-19	Sequence 19, Appl
21	304.2	60.7	4674	10	US-09-940-316B-26	Sequence 26, Appl
22	304.2	60.7	4725	10	US-09-940-316B-24	Sequence 24, Appl
23	304.2	60.7	4737	10	US-09-940-316B-30	Sequence 30, Appl
24	304.2	60.7	4767	10	US-09-940-316B-28	Sequence 28, Appl
25	304.2	60.7	4818	10	US-09-940-316B-32	Sequence 32, Appl
26	304.2	60.7	13842	9	US-09-861-289-30	Sequence 30, Appl
27	304.2	60.7	13842	9	US-09-860-846-30	Sequence 30, Appl
28	304.2	60.7	13842	10	US-09-988-384B-30	Sequence 30, Appl
29	304.2	60.7	13842	10	US-09-836-821-30	Sequence 30, Appl
30	304.2	60.7	13842	15	US-10-271-889-30	Sequence 30, Appl
31	304.2	60.7	36778	9	US-09-861-289-5	Sequence 5, Appli
32	304.2	60.7	36778	9	US-09-860-846-5	Sequence 5, Appli
33	304.2	60.7	36778	10	US-09-836-821-5	Sequence 5, Appli
34	304.2	60.7	36778	15	US-10-271-889-48	Sequence 48, Appl
35	304.2	60.7	37948	10	US-09-988-384B-5	Sequence 5, Appli
36	303.2	60.5	77536	10	US-09-940-316B-1	Sequence 1, Appli
37	303.2	60.5	64492	13	US-10-378-083-1	Sequence 1, Appli
38	302.6	60.4	6210	15	US-10-212-962-1	Sequence 1, Appli
39	302.6	60.4	85692	17	US-10-461-194-1	Sequence 1, Appli
40	300	59.9	14520	15	US-10-156-761-2885	Sequence 2885, Ap
41	299.4	59.8	11910	15	US-10-156-761-2879	Sequence 2879, Ap
42	299.4	59.8	125746	15	US-10-156-761-15102	Sequence 15102, A
43	299.4	59.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
44	297	59.3	18717	15	US-10-156-761-927	Sequence 927, App
45	297	59.3	30690	17	US-10-204-862A-1	Sequence 1, Appli

## ALIGNMENTS

### RESULT 1

US-10-156-761-2886  
; Sequence 2886, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2886  
; LENGTH: 18438  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(18438)  
US-10-156-761-2886

Query Match 65.5%; Score 328.2; DB 15; Length 18438;  
Best Local Similarity 79.5%; Pred. No. 6.3e-68;  
Matches 403; Conservative 0; Mismatches 98; Indels 6; Gaps 1;

Oy 1 TCCTTCGCTGCTGCGCTGCACATGCGGTGCGAGCCCTGCGCGGCGGCTGTC 60  
Db 12457 TCCTTCGCTGCTGCGCTGCACATGCGGTGCGAGCCCTGCGCGGCGGCTGTC 12516



Db 3613772 CGGACGGGGGGCTGGCCCGCGGACGGCCGGTGAAGCGCGTTCCGCCAGCGGGCGGACGGC 3613713  
QY 181 ACCGGCTGGTCCAGAGCGGCTGGGGCTGCTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
Db 3613712 ACGGGCTGGGGCGAGGGCGTTCGGCGTACTGCTGTGACGCGCCCTGTCGAGAGCGCGCGT 3613653  
QY 241 CGCGGGCACAGGTCTCTCGCGGTGGTCCGCGGTGGCGGTGCACTCCGACGGCGGTGCG 300  
Db 3613652 CGCGGCATGACGTGCTGGCGGTGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 3613593  
QY 301 AAGGGTTGACGGCCCGGACGGCCCGTCCAGCAGCGGTGATCCGCAAGGCACTCGCC 360  
Db 3613592 AAGGGTTGACGGCCCGGACGGGTCCGTCGAGCAGCGGTGATCCGCGAGCGCTCGCT 3613533  
QY 361 GCGCGCGGACTGTCCATCGGACGTGACGCGGTGGAGCGGACGCGCACCGGACGACC 420  
Db 3613532 TCGCGCGGGTGTTCGGCGCGGACGTGACATGCTGGAGGCGACGCGACGGGACGTCG 3613473  
QY 421 CTGGCGACCGGATCGAGGCGGCGGTGCTGGCCCACTTACGCGCAGAACCG-----G 474  
Db 3613472 CTCGGCACCGGATCGAGGCGACAGGCGTGTGGCGACGTACGCTCAGGACCGTCCGGCG 3613413  
QY 475 GAAACGCGCTGTGGCTCGGTCGGTG 501  
Db 3613412 GACCGTCCGCTGTGGCTGGGTTCGGTG 3613386

RESULT 4

US-10-156-761-412  
; Sequence 412, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 412  
; LENGTH: 18435  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(18435)  
US-10-156-761-412

Query Match 63.3%; Score 317; DB 15; Length 18435;  
Best Local Similarity 78.1%; Pred. No. 2.6e-65;  
Matches 396; Conservative 0; Mismatches 105; Indels 6; Gaps 1;  
QY 1 TCGTCTTCGCTGCTGCGCTGCACATGGCGGTGCGAGCGCCCTCGAGCGCGCGAGTGTCTCC 60  
Db 14245 TCGTCTTCGCTGCTGCGCTGCACATGGCGGTGCGAGCGCCCTCGAGCGCGCGAGTGTCTCC 14304  
QY 61 ATGGCGCTCGCGGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 120  
Db 14305 ATGGCGCTCGCGGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 14364  
QY 121 CGGACGCGGGCTGGCGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 180  
Db 14365 CGGACGCGGGCTGGCGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 14424

QY 181 ACCGGCTGGTCCAGAGCGGCTGGGGCTGCTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
Db 14425 ACCGGCTGGGCGCGAGGGCGTTCGGCACTGCTCTCTGGTGGAGCGGCTGTCCGAGCGCGCGC 14484  
QY 241 CGCGGGCACAGGTCTCTCGCGGTGGTCCGCGGTGGCGGTGCACTCCGACGGCGGTGCG 300  
Db 14485 AAGGGTTGACGGCCCGGACGGCCCGTCCAGCAGCGGTGATCCGCAAGGCACTCGCC 360  
QY 301 AAGGGTTGACGGCCCGGACGGCCCGTCCAGCAGCGGTGATCCGCAAGGCACTCGCC 360  
Db 14545 AAGGGTTGACGGCCCGGACGGGTCCGTCGAGCAGCGGTGATCCCGCAGCGCTCACT 14604  
QY 361 GCGCGCGGACTGTCCATCGGACGTGACGCGGTGGAGGCGGACGCGCACCGGACGACC 420  
Db 14605 TCGCGGGCTTACCGCGCGCGACATCGACGCGGTGGAGGCGGACGCGTACGCGGACGAGG 14664  
QY 421 CTGGCGACCGGATCGAGGCGGCGGTGCTGGCCCACTTACGCGCAGAACCG----- 474  
Db 14665 CTGGGTGACCGGATCGAGGCGGCGGTGCTGCGCACGTACGCGCAGGACGCGCGCCGCA 14724  
QY 475 GAAACGCGCTGTGGCTCGGTCGGTG 501  
Db 14725 GACAGCGCTGTGGCTGGGTTCGGTG 14751

RESULT 5

US-10-156-761-15103  
; Sequence 15103, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 15103  
; LENGTH: 100000  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-15103

Query Match 63.3%; Score 317; DB 15; Length 100000;  
Best Local Similarity 78.1%; Pred. No. 1.8e-65;  
Matches 396; Conservative 0; Mismatches 105; Indels 6; Gaps 1;  
QY 1 TCGTCTTCGCTGCTGCGCTGCACATGGCGGTGCGAGCGCCCTCGAGCGCGCGAGTGTCTCC 60  
Db 64768 TCGTCTTCGCTGCTGCGCTGCACATGGCGGTGCGAGCGCCCTCGTAAACGCGGAGTGTCTCC 64827  
QY 61 ATGGCGCTCGCGGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 120  
Db 64828 ATGGCGCTCGCGGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 64887  
QY 121 CGGACGCGGGCTGGCGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 180  
Db 64888 CGGACGCGGGCTGGCGCGCGGTGATGATGGGCAAGGTTCGCGCAAGGTTCGCGCAAGGTTCGCG 64947  
QY 181 ACCGGCTGGTCCAGAGCGGCTGGCGCGGTGCTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240  
Db 64948 ACCGGCTGGGCGGAGGCGCTGCGCATGCTCTCTGGTGGAGCGGCTGTCCGACGCGCGCGC 65007



```
QY 301 AACGGCTTGACGGCCCGAAGCGCCCTCCAGCAGCGCGTGATCCCGAAGGCACCTGGCC 360
Db 916 AACGGCTTGACGGCCCGAAGCGCCCTCCAGCAGCGCGTGATCCCGAAGGCACCTGGCC 975
QY 361 GCCCGCGGACTGTCCACATCGACGCTCGACGCGTGGAGCGGCGACCGCACCGGACGACC 420
Db 976 GGGCGCCGGCTGTCCGCGGAGGAGTGCAGCGCTGCGAAGCACACGCGCACGCGGACCGAGC 1035
QY 421 CTGGGCGACCGGATCGAGGCGGAGCGCTGTGGCCACCTACGCGCCAGAACCGGGAACG 480
Db 1036 CTCGGCGACCGGATCGAGGCGGAGCGCTGTGGCGACCTATGGAAGAGCGCGGAACG 1095
QY 481 CCGCTGTGGCTCGGCTCGGTG 501
Db 1096 CCGCTGTGGCTCGGCTCGGTG 1116

RESULT 8
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; PRIOR FILING DATE: 2001-08-27
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

Query Match 62.8%; Score 314.4; DB 10; Length 77536;
Best Local Similarity 76.8%; Pred. No. 7.6e-65;
Matches 384; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 TCGTCTTGCTGGTGGCTGACATGGCGGTGTCAGGCCCTCGAGCGCGCGGAGTGCTCC 60
Db 42592 TCGTCTTGCTGGTGGCTGACATGGCGGTGTCAGGCCCTCGAGCGCGCGGAGTGCTCC 42533
QY 61 ATGGGCGCTCGCGGCGGCGTGTGATGGGCGACGCTGCGAGACGTTCTCGAGTTCTCG 120
Db 42532 CTGGGCGCTGGTGGCGGCTGTACCGTCAATGGCGAACACCGAGAGTCTCGTGGATCTCC 42473
QY 121 CGGCGACGCGGGCTGGCCCCCGACGCGCGCTGCAAGCGGTTCCCGACCGCGCGGACGCG 180
Db 42472 CGGCGACGCGGACTCTCTCCCGGACGACGCGTGCAGGCGTTCCCGACCGCGCGGACGCG 42413
QY 181 ACCGGTGTGTCGAGGCGGCTGGGCTGTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240
Db 42412 ACCGGTGTGTCGAGGCGGCTGGGCTGTCTCTGGTGGAGCGGCTGTCCGAGCGCGGCGC 42353
QY 241 CGCGGCGACCAAGTTCCTCGCCGTGGTCCGCGGCTCGCGGCTCAACTCCGACGCGCGCTCG 300

Query Match 61.9%; Score 310.2; DB 15; Length 10056;
Best Local Similarity 76.4%; Pred. No. 1.1e-63;
Matches 381; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 1 TCGTCTTGCTGGTGGCTGACATGGCGGTGTCAGGCCCTCGAGCGCGCGGAGTGCTCC 60
Db 622 TCGTCTTGCTGGTGGCTGACATGGCGGTGTCAGGCCCTCGAGCGCGCGGAGTGCTCC 681
QY 61 ATGGGCGCTCGCGGCGGCGTGTGATGGGCGACGCTCGAGACGTTCTCGTGGATCTCG 120
Db 682 ATGGGCGTTGGCGGCGGCGTGTGAGGTGATGACACGCTTACCGCGTTCTGGAGTTCAGT 741
QY 121 CGGCGACGCGGGCTGGCCCCCGACGCGCGCTGCAAGGCGTTCCCGACCGCGCGGACGCG 180
Db 742 CGCCAGCGGGCTGGCCCCCGACGCGCGTGCAGGCGTTTGGTGGCGCGCGGACGCG 801
QY 181 ACCGGTGTGTCGAGGCGGCTCGGGCTGTCTCTGGTGGAGCGGCTGTCCGAGGCTCAGCGT 240
Db 802 ACCGGTGGCGGCGGAGGCGGCTCGGCATGTCTCTGGTTCGAGCGGCTGTCCGACGCGGCGC 861
QY 241 CGCGGCGACCAAGTTCCTCGCCGTGGTTCGCGGCTGGCGGCTCAACTCCGACGCGCGCTCG 300
Db 862 AACCGGCGACGAGTTCCTTGGGTGGTACCGGGGAGCGCGGTGAACACGAGGACGCTGCGTCC 921
```

QY 301 AACGCTTGAAGCGCCCGAAGCGCGCTCCAGCAGCGCGTGTATCCGGAAGCACTGGCC 360  
 Db 922 AACGCTTGAAGCGCCCGAAGCGCGCTCCAGCAGCGCGTGTATCCGGAAGCACTGGCC 981  
 QY 361 GCCGCGGAGCTTCCACATCGGACGCTCGACGCGGTGGAGCGCGCAGCGCAACCGGACGCC 420  
 Db 982 TCGGCGGAGCTTCTCCCGCCAGATCGACGCTGTGAGCGCGCAGCGCAACCGGACGCC 1041  
 QY 421 CTGGGCGAGCCCGATCGAGCGCGGCGCTGTGGCCACCTAGCGCGCAACCGGAGAACG 480  
 Db 1042 CTGGGCGAGCCCGATCGAGCGCGGCGCTGTGGCCACCTAGCGCGCGGAGCGGACGAG 1101  
 QY 481 CCGCTGTGCTCGGCTCGG 499  
 Db 1102 GACCGCGCGCTGTGGTTGG 1120

RESULT 10  
 US-10-156-761-414  
 ; Sequence 414, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 414  
 ; LENGTH: 10692  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(10692)  
 US-10-156-761-414

Query Match 61.8%; Score 309.6; DB 15; Length 10692;  
 Best Local Similarity 77.3%; Pred. No. 1.5e-63;  
 Matches 391; Conservative 0; Mismatches 109; Indels 6; Gaps 1;  
 QY 1 TCGTCTTCGCTGTCGCGCTGCACATGCGCGTGCAGCGCGCTGTGAGCGCGGAGTGTCTCC 60  
 Db 619 TCGTCTTCGCTGTCGCGCTGCACATGCGCGTGCAGCGCGCTGTGAGCGCGGAGTGTCTCC 678  
 QY 61 ATGCGCTTCGCGCGCGCGCTGTGATGATGGCAGCGTGCAGAGCTTCCTCGAGTTCTCG 120  
 Db 679 CTGCGCGTTCGCGCGCGCGCTGTGATGATGGCAGCGTGCAGAGCTTCCTCGAGTTCTCG 738  
 QY 121 CGGCGAGCGCGCGCTGGCGCCCGAAGCGCGCTGTGCAAGGCGTTTCGCGCGCGCGGACGGC 180  
 Db 739 CGCAGCGCGGACTGGGACACCGACGCGCGCTGTGCAAGGCGTTTCGCGCGCGCGGACGGC 798  
 QY 181 ACCGCGTGTGTCGAGAGCGCTGCGGCTGTCTCTGTGAGCGGCTGTGCGAGGCTCAGGCT 240  
 Db 799 ACCGCGTGTGTCGAGAGCGCTGCGGCTGTCTCTGTGAGCGGCTGTGCGAGGCTCAGGCT 858  
 QY 241 CGCGGGACACAGGCTCCTCGCGGTGTGTCGCGGTGCGGCTGCGGCTCACTCCGAGCGGCGCTCG 300  
 Db 859 AACGGGACGAGGCTCTTTCGCGGTGTGTCGCGGTGCGGCTGCGGCTCACTCCGAGCGGCGCTCG 918  
 QY 301 AACGCTTGAAGCGCCCGAAGCGCGCTCCAGCAGCGCGTGTATCCGGAAGCACTGGCC 360

Db 919 AACGCTTGAAGCGCCCGAAGCGCGCTCCAGCAGCGCGTGTATCCGCGAGCGCTGGCC 978  
 QY 361 GCCGCGGAGCTTCCACATCGGACGCTCGACGCGGTGGAGCGCGCAGCGCAACCGGACGCC 420  
 Db 979 GCCGCGGAGCTTCCACATCGGACGCTCGACGCGGTGGAGCGCGCAGCGCAACCGGACGCC 1038  
 QY 421 CTGGGCGAGCCCGATCGAGCGCGGCGCTGTGGCCACCTAGCGCGCAACCGG-----G 474  
 Db 1039 CTGGGCGAGCCCGATCGAGCGCGGCGCTGTGGCCACCTAGCGCGCGGCGCTGGCGGAC 1098  
 QY 475 GAAAAGCGCGCTGTGCTCGGCTCGGTT 500  
 Db 1099 GACAAGCGCGCTGTGCTGGCTCGAT 1124

RESULT 11  
 US-10-156-761-2882  
 ; Sequence 2882, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 2882  
 ; LENGTH: 10839  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(10839)  
 US-10-156-761-2882

Query Match 61.7%; Score 309; DB 15; Length 10839;  
 Best Local Similarity 77.1%; Pred. No. 2.1e-63;  
 Matches 391; Conservative 0; Mismatches 110; Indels 6; Gaps 1;  
 QY 1 TCGTCTTCGCTGTCGCGCTGCACATGCGCGTGCAGCGCGCTGTGAGCGCGGAGTGTCTCC 60  
 Db 5203 TCCTCTCTCGTGTGTGCTGACCTGCGCAGGGCAGCGCGCTGCGGGGCGCGAGTGGCAG 5262  
 QY 61 ATGCGCTTCGCGCGCGCGCTGTGATGATGGCAGCGTGCAGAGCTTCCTCGAGTTCTCG 120  
 Db 5263 CTGCGCTGTGTCGCGCGCTGTGACCGTATGCGCCACGCGCGGAGCTTCCTGTCAGTTCTCG 5322  
 QY 121 CGGCGAGCGCGCGCTGGCGCCCGAAGCGCGTTCGCGCGCGCGGACGGCAGCGGACGGC 180  
 Db 5323 CGGCGAGCGCGCTTCGCGCGGAGACCGCGCGTGTGCAAGGCGTTTCGCGCGCGGTCGCGAGCGG 5382  
 QY 181 ACCGCGTGTGTCGAGAGCGCTGCGGCTGTCTCTGTGAGCGGCTGTGCGAGGCTCAGGCT 240  
 Db 5383 ACCGCGTGTGTCGAGAGCGCTGCGGCTGTCTCTGTGAGCGGCTGTGCGAGGCTCAGGCT 300  
 QY 241 CGCGGGACACAGGCTCCTCGCGGTGTGTCGCGGTGCGGCTCACTCCGAGCGGCGCTCG 360  
 Db 5443 GACGCGGCTCGGCTGTGTCGCGGCTGTGTCGCGGCTTCGCGCGGCTTCGCGAGCGGCTCGGCT 5502  
 QY 301 AACGCTTGAAGCGCCCGAAGCGCGCTCCAGCAGCGCGTGTATCCGGAAGCACTGGCC 360  
 Db 5503 AACGCTTGAAGCGCCCGAAGCGCGCTTCGCGAGCGGCTGTATCCGCGAGCGGCTCGCT 5562



```
QY 361 GCGCGCGACTGTCCACATCGGACGTGCGACCGGTGAGCGCGCACCGCACCGGACGACC 420
Db 5563 TCGCGCGGGTGTTCGGCGCGCGACGTGGACATGTTGAGCGCGCACCGGACGACGTG 5622
QY 421 CTGGCGGACCGGATCAGGCGCGGAGGCGTCTGCGCCACCTTACGGCCAGAACCG-----G 474
Db 5623 CTGGCGGACCGGATCAGGCGCACAGGCGTCTGCTGGCGACGTACGGTCAGGACCGTCCGGCG 5682
QY 475 GAAACGCGCTGTGGCTCGGGTCGGTG 501
Db 5683 GACCGTCCGCTGTGGCTGGGTTCGGTG 5709

RESULT 12
US-10-461-194-2
; Sequence 2, Application US/10461194
; Publication No. US20040077058A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schitmer, Andreas
; APPLICANT: McDaniell, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GEUDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009700
; CURRENT APPLICATION NUMBER: US/10/461,194
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 86941
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-10-461-194-2

Query Match 61.7%; Score 309; DB 17; Length 86941;
Best Local Similarity 76.0%; Pred. No. 1.3e-63;
Matches 381; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGGTGGCGCTGCACATGCGCGTGCAGGCGCCCTGCAGCGCGGAGTGTCC 60
Db 31457 TCGTCTTCGCTGGTGGCGCTGCACATGCGCGTGCAGGCGCGTTCAGGCGGTGCGTACGGCGGAAATGCTCG 31516
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTATGGGACGCTGAGACGCTTCGTGAGTTCCTCG 120
Db 31517 CTGGCGCTCGCGCGCGGTGTCGCGGTGATGACGACGCGCGACGCTTCATGAGTTCCTCC 31576
QY 121 CGGACGCGCGGCTGCGCCCGCGCGCTGCAAGGCGTTCGCGACGCGCGCGACCGGC 180
Db 31577 GCGACGCGCGGCTCGCCCGCGCGATGCGCGCTGATGAGTGTTCGCGAGGCGCGCGACCGGC 31636
QY 181 ACCGCTGTGTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240
Db 31637 ACCAAGTGTTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240
QY 181 ACCGCTGTGTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240
Db 31637 ACCAAGTGTTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240

QY 241 CGCGGCGACGAGGTCTCTCGCGTGGTTCGCGGCTCGGCTCAACTCCGACGGCGGTGCG 300
Db 31697 ATGGGCGATGAGGTGCTCGCGGTGGTTCGCGTGGCACGGGCTGAAACGAGGACGGCGGAGC 31756
QY 301 AACGGCTTGAACGCGCCCGAACACGCGCCCTCCAGACGCGGTGATTCGCAAGGCACTGGCC 360
Db 31757 AACGGCTTGAACGCGCCCGAACACGCGCCCTCCAGGAAACGGGTGATTCGCGCAGGCGCTGGCG 31816
QY 361 GCGCGCGGATGTCCACATCGGACGTGAGCGGTGGAGCGCACGSCACCGGACGACG 420
Db 31817 AACGCGCGGCTGACGCGGTGCGCGGTGAGCGCGTTCGAGCGCACCGGACCGGACGAGT 31876
QY 421 CTGGCGGACCGGATCGAGGCGCGGCGCTCTGCGCCACCTTACGCGCCAGAACCGGAAACG 480
Db 31877 CTGGCGGACCGGATCGAGGCGCGGCGCTCTGCGCCACCTTACGCGCAGGAGCGCGCGGAG 31936
QY 481 CGCGTGTGGCTCGGGTCGGTG 501
Db 31937 GGTGAGCGGCTGTGGCTGGGG 31957

RESULT 13
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 61.4%; Score 307.4; DB 9; Length 15872;
Best Local Similarity 76.9%; Pred. No. 4.5e-63;
Matches 390; Conservative 0; Mismatches 111; Indels 6; Gaps 1;

QY 1 TCGTCTTCGCTGGTGGCGCTGCACATGCGCGTGCAGGCGCCCTGCAGCGCGGAGTGTCC 60
Db 9017 TCCTCTCGCTGGTGGCGCTGCATCTGGCGGTACGGCGCTGCGCACGGCGAGTGGCGG 9076
QY 61 ATGGCGCTCGCGCGCGCGGTGATGGTATGGGACGCTGCGAGACGCTTCGTGAGTTCCTCG 120
Db 9077 CTGCGCTGCGCGCGCGGTGCGGTGATGCGCGCTGATGCGCGATCCGCGCGCGTTCGTGAGTTCCTCC 9136
QY 121 CGGACGCGCGGCTGCGCCCGCGCGCTGCAAGGCGTTCGCGACGCGCGCGACCGGC 180
Db 9137 CGGACGCGCGGCTGCGCCCGCGCGCTGCAAGGCGTTCGCGCGCGCGCGCGCGCGC 9196
QY 181 ACCGCTGTGTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240
Db 9197 ACCGCTGTGTCGAGGCGGTGCGGCTGCTCTCTGTTGGAGCGGCTTCGAGGCTCAGCGT 240
QY 241 CGCGGCGACGAGGTCTCTCGCGTGGTTCGCGGCTGCGCGGTCAACTCCGACGCGCGGTGCG 300
Db 9257 GCGGCGCACACGCTCTCTCGCGTGGTTCGCGCGCTGCAAGGCGTTCGCGCGCGCGCGCGC 9316
QY 301 AACGCTTGAACGCGCCCGAACGCGCGCTTCGAGCAGCGGCTGATCCGCAAGGCACTGGCC 360
Db 9317 AACGCTTGAACGCGCCCGAACGCGCGCTTCGAGCAGCGGCTGATCCGCAAGGCACTGGCC 360
QY 361 GCGCGCGGACTGTCCACATCGGACGCTGCGACGCGGTGGAGCGCGCACGCGCACCGGACGACC 420
```

Db	9377	GACGCCGGGCTGTCTCCCGGAGGACGTGTGACGGCGTCTGAGGCGCACCGGCACCGCGC	9436
QY	421	CTGGGGGACCCGATCCAGAGCCCGAGGGCGCTGCTGGCCACCTACGGCCAGAACCC	474
Db	9437	CTCGGGGACCCCATCCAGAGCCGGGGCGCTGTCTCGCCGCTCCGAGGAGAACCGTTCCGCGC	9496
QY	475	GAACACCGCGCTGTGGCTCGGGTCGGTG	501
Db	9497	GACCACCGCTGTGGCTCGGCTCGCTG	9523

RESULT 14

US-09-860-846-1

; Sequence 1, Application US/09860846

; Patent No. US20020164742A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; CURRENT FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 15872

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-860-846-1

Query Match	61.4%	Score 307.4	DB 9	Length 15872
Best Local Similarity	76.9%	Pred. No. 4.5e-63		
Matches 390	Conservative 0	Mismatches 111	Indels 6	Gaps 1
QY	1	TCGTTCTTCGCTGCTCGCGCTGCACATGGCGGTGCAGGCCCTTCAGCGCGCGAGTGTCTCC	60	
DB	9017	TCCTCTCGCTGCTGCGCTGCATCTGCGGTACCGGCGCTTGGCGCACGCGAGTGTGGGG	9076	
QY	61	ATGGCGCTCGCGGCGCGGTGCATGCTGATGAGGCAAGTTCGAGACGTTCTGTCGAGTTCTCG	120	
DB	9077	CTCGCGCTGGCGGGGGGGTGGCGTGAATGCCGATTCGGCGCGGTTCTGTGAGTTCTCC	9136	
QY	121	CGGACGCGGCGCTGGCCCCCGACGGCCCGCTTCAAGCGGTTTCGCCGACGGCGGACGCGC	180	
DB	9137	CGGCGAGAGGGGCTGGCGCCGACGGCCGCTGCAAGGGTTCTCGGCGCGCGCGACGCGC	9196	
QY	181	ACCGGCTGTTCCGAGGGCGTCCGGGTGTTCTCTGTGTGAGCGGCTGTTCGAGGCTCAGCGT	240	
DB	9197	ACCGGCTGGCGCGAGGGCGTTCGGCGTGTCTGTTCTGTGAGCGGCTGTTCGACGCGCGCGC	9256	
QY	241	CGGGGACACGAGTCTTCGCGGTGTTCGGCGGTTCGGCGGTCAACTCCGACGGCGCGCTCG	300	
DB	9257	GCGGGGCAACAGTCTCTCGGCGTGTTCACCGGCTCAACCGGCTCAACGAGGACGTTGCTCC	9316	
QY	301	AACGGCTTACAGGCGCCCGAAGCGCCCGTCCAGACGAGCGGCTATCCGCAAGGCACTGGCC	360	
DB	9317	AACGGGCTACCGCGGCCAACCGCCGACGCCACAGCAACGCGTATCGCCGAGGCGCTGCC	9376	
QY	361	GCGCGCGGACTGTTCACATCGGACGTTCCAGCGGTTGGAGGGCGCACGACCGGGACGACC	420	
DB	9377	GAGCGGGGCTGTCCCGGAGGACGTGTGACGCGGTCGTGAGGCGCACGCGACCGGACCCGG	9436	
QY	421	CTGGGCGACCCGATCGAGGCGCGAGGCGTGTCTGGGCCCTACGGCCAGAACCC-----GG	474	
DB	9437	CTCGGCGACCCCATCGAGGCGGGGGCGTGTCTCGCGGCTTCGGACGGAACCGTTCCGCGC	9496	
QY	475	GAAACGCGCGTGTGGTTCGGTTCGGTG	501	

```

Db      9497  GACCACCGCTGTGGCTCGCTGCTG 9523

RESULT 15
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

```

Query Match	61.4%;	Score 307.4;	DB 10;	Length 15872;
Best Local Similarity	76.9%;	Pred. No. 4.5e-63;		
Matches 390;	Conservative 0;	Mismatches 111;	Indels 6;	Gaps 1;
QY	1	TCGCTTCGCTGCTCGCGCTGCACATCGCGGTGCAGCGCCCTGACGCGCGCGAGTGTCTCC	60	
Db	9017	TCCTCCTCGCTGCTGCGCTGCATCTGGCGGTACGGCGCTGCGGCACGCGAGTGTCTCC	9076	
QY	61	ATGGCGCTTCGCGGGCGCGGTGATGGTGAATGGGCACGGTTCGAGACGTTCTGTCAGTTCCTG	120	
Db	9077	CTCGCGCTTCGCGGGGTGGCGGTGATGSCCGATCCGGCGCGTTCGTGGAGTTCCTCC	9136	
QY	121	CGGCAGCGGGGCTGGCCCCCGACGCGCGCTGCAGAGGCGTTCGCCGAGCGCGCGACGCGC	180	
Db	9137	CGGCAGAGGGGCTGGCCCGCCGACGCGCGCTGCAGGCGTTCCTCGCGCGCGCCGACGCGC	9196	
QY	181	ACCGGCTGTCGAGGGGCTCGGGCTGCTCTCTGCTGGAGCGGCTGTCGAGGCTCAGCGT	240	
Db	9197	ACCGGCTGCGCGAGGGCTCGCGTGTCTCTCTGGAGCGGCTGTCGACGCGCGCGC	9256	
QY	241	CGCGGGCACACAGTCTCTCGCCGTGGTTCGCGGGTTCGCGGTCAAATCGACGGCGCGTGG	300	
Db	9257	GCGGGGCACACGTCTCTCGGCTTGTTCACCGGCACCGCGGTCAACAGGACGGTGCCTCC	9316	
QY	301	AACGGGTTGACGGCCCCGAAACGGCCCCGTCCAGCAGCGCGCTGATCCGCAAGGCACTGGCC	360	
Db	9317	AACGGGCTGACCGCGCCCAACCGGCCACGCCAGCAACCGCTCATCGCCGAGGCGCTGCC	9376	
QY	361	GCGCGGACTCTGCACATCGGACGCTGCACGGGTGAGGGCGCAGCGCACCGGGACGACC	420	
Db	9377	GACCCGGGCTGTCCCGAGAGACCTGTGACGCGGTCGAGCGGCACGCCACCGGCACCCGG	9436	
QY	421	CTGGGCGACCCGATCGAGSCCGAGCGCTGCTGGCCACTTACGGCCAGAACG-----GG	474	
Db	9437	CTCGGCGACCCCATCGAGSCCGGGCGTGTCTCGCGGCTCCGGACGGAAACGTTCTCGGC	9496	
QY	475	GAAACCGCGCTGTGGCTCGGGTCGGTG	501	
Db	9497	GACCAACCGCTGTGGCTCGGCTCGCTG	9523	

Search completed: June 17, 2004, 03:37:02  
Job time : 311.667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      June 16, 2004, 23:18:23 ; Search time 1960 Seconds
              (without alignments)
              7633.139 Million cell updates/sec

Title:      US-10-042-665A-3_COPY_8434_8934
Perfect score: 501
Sequence:    1 tcgtttcgtcggtcgcgct.....cgcgtggtcgggtcggtg 501

Scoring table:  IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:      27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:  55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :
1:  em_estba:*
2:  em_estum:*
3:  em_estin:*
4:  em_estmu:*
5:  em_estov:*
6:  em_estpl:*
7:  em_estro:*
8:  em_hrc:*
9:  gb_estl:*
10: gb_est2:*

```

[illegible]

BZ561977 753 bp DNA linear GSS 17-DEC-2002  
pac82-164\_3530.x1 pac82-164 Pseudomonas aeruginosa genomic clone  
pac82-164\_3530, genomic survey sequence.  
BZ561977  
BZ561977.1 GI:27183081  
GSS.  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 753)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	141.6	28.3	753	28	BZ561977	BZ561977 pacs2-164
2	116	23.2	871	28	BZ566693	pacs2-164
C	104.2	20.8	625	12	BM091001	BM091001 I921a10.Y
3	103.6	20.7	865	28	BZ566683	BZ566683 pacs2-164

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@imagine.wustl.edu)  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 485.

# FEATURES

Location/Qualifiers  
1..625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="Fetal Pancreas (4 Pooled Donors, 18 - 20 weeks, Stragene #738023)"  
/dev\_stage="Fetal Pancreas"  
/clone\_lib="Human Fetal Pancreas 1B"  
/note="Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110 E-mail: hinoue@imagine.wustl.edu Tel: 314-362-1916, Fax: 314-747-2692."

# ORIGIN

Query Match 20.8%; Score 104.2; DB 12; Length 625;  
Best Local Similarity 52.3%; Pred. No. 0.00014;  
Matches 260; Conservative 0; Mismatches 228; Indels 9; Gaps 1;

```

Qy 1 TCGTCTTCGCTGTCGCGCTGCACATGGCGGTGAGGCCCTGACGGCGGGGAGTGTCC 60
Db 604 TCTTCCAGCTGTAGTGGCCCTGCAGACGCTTACCAGGCCATCCACAGCGGCATGCGCT 545
Qy 61 ATGCGCTGCGGGCGGCGTGTATGTATGGCAGCGTGCAGAGCTTCTCGAGTTCG 120
Db 544 GCGGCCATCTGTGGGGGATCAAGTCTCTGTGAAGCCACACTCGTGCAGTTCG 485
Qy 121 CGGCAGCGCGGCTGCGCCCGACGCGCGCTGCAAGCGTTCGCCGACGCGCGGACGCG 180
Db 484 AGGCTGGGATGCTCAGCCCGAGGCGACCTGCAAGGCTTCGACAGCGGGAATGG 425
Qy 181 ACCGGCTGTGCGAGGGCGTGGGCTGTCTCTGTGTGAGCGGTGTGCGAGGCTCAGGT 240
Db 424 TACTGCGGCTCGAGGGGTGTGTGTGCGGCTCTCTGTGACCAAGAGTCTCCGCGCGCGG 365
Qy 241 CGCGGACACAGGTCCTCGCGTGTGCGGCGGTGCGGCGGTCACTCCGACGCGCGCTG 300
Db 364 GTGTAGGCCACATCTCTGAACGCGGACCAATACAGATGGCTTCAAGGA-----G 314
Qy 301 AACGGCTTGAAGGCGCCCGGAAAGCGGCCCTGCCAGCAGCGCGTGTATCCGCAAGGCACTGGCC 360
Db 313 CAAGGCGTGACTTCCCTCAGGGGATATCCAGGAGCAGCTCATCCGCTGTTGTACCAG 254
Qy 361 GCGCGCGAGTGTTCACATCGAGCTCGACGGGTGTGAGGGCGCAGCGCACCGGGAGCACC 420
Db 253 TCGGCGGAGTGGCCCTTGAGTTCATTGATATACATCGAAGCCCGACGCGACAGCCCAAG 194
Qy 421 CTGCGCAGCCGATCGAGCGCGGCGTGTGTGCGCCACTACGGCCAGAACCGGGAACG 480
Db 193 GTGGCGGACCCCGAGAGCTGAATGGCATACCCGAGCCCTGTGTGCGCCACCGCCAGGAG 134
Qy 481 CCGCTGTGGCTCGGTC 497
Db 133 CCGCTGTCTATCGGCTC 117

```

# RESULT 4

BZ566683  
LOCUS BZ566683 865 bp DNA linear GSS 17-DEC-2002  
DEFINITION pacs2-164\_6520.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_6520, genomic survey sequence.  
ACCESSION BZ566683  
VERSION BZ566683.1 GI:27196700  
KEYWORDS GSS.

# SOURCE

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

# REFERENCE

AUTHORS 1 (bases 1 to 865)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
CONTACT: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
CLASS: shotgun.

# JOURNAL

COMMENT Location/Qualifiers  
1..865  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164\_6520"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun  
library."

# FEATURES

source  
1..865  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164\_6520"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun  
library."

# ORIGIN

Query Match 20.7%; Score 103.6; DB 28; Length 865;  
Best Local Similarity 63.0%; Pred. No. 0.00017;  
Matches 160; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
Qy 248 ACCAGGTCTTCGCGGTGTGTCGCGGTCGCGGTCACTCCGACGCGGCGCTCGAACGCT 307  
Db 92 ACCGGGTATCGCCGCTTCTGCTGGCTCGCGGTCACCATGACGCGCCCTCGCGGGGCG 151  
Qy 308 TGACGGGCCCCGAAAGCGCGCTCCAGCAGCGCGTGCATCCGCAAGGCACTGGCGCGCGCG 367  
Db 152 TCACGGTACCCACAGGACCGGCCAGAGGGGCTTGTTCGGAGGCGTGGCCACGCCA 211  
Qy 368 GACTGTCCATCGAGCGTGCACGCGGTGTGAGGCGCACGCGCACCGGAGACCCCTGGGCG 427  
Db 212 GCTGTGCGCGCGCGGTGTGCGGCTATGTCGAGGCCCATGTCGACACGTCGTCGCGG 271  
Qy 428 ACCGATCGAGCGCGGCTGTGCGCACCTACGGCCAGNACCGGGAACGCCGCTGT 487  
Db 272 ATCCGATCGAGTTGCGGCGCTTGCCACGCTATCGAGGTACTGAGCGGTGCCCCCTGG 331  
Qy 488 GCCTCGGGTCCGCTG 501  
Db 332 CGTGGCATCGGTG 345

# RESULT 5

BZ577421  
LOCUS BZ577421 1303 bp DNA linear GSS 17-DEC-2002  
DEFINITION msh2\_5402.x1 msh Pseudomonas aeruginosa genomic clone msh2\_5402,  
genomic survey sequence.

# ACCESSION

VERSION BZ577421  
GI:27212482

# KEYWORDS

GSS.

# SOURCE

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

# REFERENCE

1 (bases 1 to 1303)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press



TITLE Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
*Pseudomonas aeruginosa* library  
 JOURNAL J. Bacteriol. (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 20622216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

## FEATURES

source  
 1. .979  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164 7558"  
 /clone\_lib="pacs2-164"  
 /notes="Clinical isolate 2-164 Whole genomic shotgun  
 library."

## ORIGIN

Query Match 19.6%; Score 98; DB 28; Length 979;  
 Best Local Similarity 58.1%; Pred. No. 0.00093;  
 Matches 173; Conservative 0; Mismatches 125; Indels 0; Gaps 0;  
 QY 1 TCGTTCGTGTCGCGCTGCATCGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCC 60  
 Db TCGTTCGTGTCGCGCTGCATCGCGGTGCGAGCCCTGCGAGCGCGGAGTGTCC 330  
 QY 61 ATGGCGCTCCGGCGCGGTGATGTCATGCGACGCTCGAGAGCTTCGTCGAGTTC 120  
 Db ATTGCTATGTCGCGCGGTGATGTCATGCGACGCTCGAGAGCTTCGTCGAGTTC 270  
 QY 121 CGGCGAGCGCGGCTGCGCCCGCGCGCTGCGAAGCGGTTCGCCGACGGCGCGG 180  
 Db GATCCCGTGGCTATCGGCGGACGGCACCTGCGAAGACCTTCGACGATGTCGCG 210  
 QY 181 ACCGCTGTCGCGGCGTTCGCGCTGCTCCTGTCGAGCGGCTGTCGAGGCTCAG 240  
 Db TACGCGCGCGCGAGGGTTGCGGCTGTCGATCTCAAGCGCGCGGACCGGCGCG 150  
 QY 241 CGCGGCGACAGGTCCTCGCGCTGTCGCGCGCTGTCGCGGCTCACTCGACGCG 298  
 Db GACGCGACCGGGTCATCGCGCTTCTGTCGGCTCGCGGCTCAACCATGACGGCG 92

## RESULT 8

BI590480 444 bp mRNA linear EST 06-SEP-2001  
 LOCUS RH04107.5prime RH Drosophila melanogaster normalized Head pFlc-1  
 DEFINITION Drosophila melanogaster cDNA clone RH04107 5 similar to  
 BcDNA:GH07626; FBan0003523 GO:[fatty-acid synthase (GO:0004312)]  
 located on: 2L 23D1-23D1.; 08/16/2001, mRNA sequence.

## ACCESSION

VERSION BI590480.1 GI:15481902  
 SOURCE EST.

## KEYWORDS

ORGANISM Drosophila melanogaster (fruit fly)

LOCUS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,  
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,  
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,  
 Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,  
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and  
 Rubin, G.M.

TITLE BDGP/HMI RH Drosophila EST Project

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: http://www.fruitfly.org/EST. est@fruitfly.berkeley.edu  
 hit genomic AE003581: arm:2L [2747932,3054367]  
 estimated-cyto:23B3-23D3: 08/16/2001  
 Plate: RH.41 row: A column: 7  
 High quality sequence stop: 423.

## FEATURES

source  
 1. .444  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="RH04107"  
 /sex="male and female"  
 /dev\_stage="Adult"  
 /lab\_host="DH5-alpha Tona"  
 /clone\_lib="RH Drosophila melanogaster normalized Head  
 pFlc-1"  
 /notes="Organ: head; Vector: pFlc1; site 1: XhoI; Site 2:  
 BamHI; Library was kindly generated by Piero Carninci at  
 the RIKEN. The library was normalized and excised using  
 Cre recombinase. Plasmid cDNA library."

## ORIGIN

Query Match 17.6%; Score 88.4; DB 12; Length 444;  
 Best Local Similarity 53.5%; Pred. No. 0.016;  
 Matches 213; Conservative 0; Mismatches 176; Indels 9; Gaps 1;  
 QY 103 ACGTTCGTGAGTTCCTCGGCGAGCGGGTGGCCCCCGGCGCGCTGCAAGCGGTC 162  
 Db ATGTGCTGCTGAGTTCCTCGGCGAGCGGGTGGCCCCCGGCGCGCTGCAAGCGGTC 66  
 QY 163 GCCGACGGCGGACGCGCGCTGTCGAGGGCGTTCGGGCTGCTCTCTGTGGAGCGG 222  
 Db GATGAGTATGCAATGATGATGTCGCTTCCATCTCAATGTGCGCACCAATACGAT 126  
 QY 223 CTGTCCGAGGTCACAGCTCGCGGCGACGAGTCTCTCGCGGTTCGCGGGTCGCG 282  
 Db ACCTCTGCGCGCGCGTGTGTATGCTTCCATCTCAATGTGCGCACCAATACGAT 186  
 QY 283 AACTCCGACGCGCTGCAAGCTTGAAGCGGCGGCGCGGCGCGGCGCGCGCGG 342  
 Db TTCAAGGA-----GCAGGGCATCACATACCTATTGGCAAGATGCAAAATCG 237  
 QY 343 ATCCGCAAGGCACTGGCGCGCGCGGCTGTCATCGACGTCGACGCGGTGAGGCG 402  
 Db ATCCGCGAGACCTAGAGGAGATGGTCTTAACCCCGCGATGTTTACGTGAGGCA 297  
 QY 403 CACGCGACCGGAGACCCCTCGGCGACCCGATCGAGGCGGCGGCTCTGGCCACCTAC 462  
 Db CAGGTACCGGAAACCAAGTGGCGGATCCCGCAGAGGTGAACCTATCACTGACTTCT 357  
 QY 463 GCGCAGAACCGGGAACCGCCCTGTGGCTCGGCTCGGT 500  
 Db TGCAAGGACCGTACGACCCCTCTGTTGATCGGTCGGT 395

## RESULT 9

BI570498 502 bp mRNA linear EST 06-SEP-2001  
 LOCUS RH02843.5prime RH Drosophila melanogaster normalized Head pFlc-1  
 DEFINITION Drosophila melanogaster cDNA clone RH02843 5 similar to  
 BcDNA:GH07626; FBan0003523 GO:[fatty-acid synthase (GO:0004312)]  
 located on: 2L 23D1-23D1.; 07/26/2001, mRNA sequence.

## ACCESSION

VERSION BI570498.1 GI:15461920

## KEYWORDS

ORGANISM Drosophila melanogaster (fruit fly)

LOCUS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;





```

QY 403 CACGGCACCGGAGACACCTCGGGGACCCGATCGAGGCCGAGGCGCTGCTGGCCACCTAC 462
Db 298 CACGGTACCGGAACCAAGTGGGCGATCCCGAGGAGGTGAATCTATCACTGACTTCTTC 357
QY 463 GCGCAGACCGGGAACCGCGCTGTGGCTCGGTC 497
Db 358 TGCAGAGACCGGTACGACCCCTCTGTGTATCGGATC 392

RESULT 11
LOCUS BE299394/1 766 bp mRNA linear EST 20-JUL-2000
DEFINITION 601118872F1 NTH_MGC_17 Homo sapiens cDNA clone IMAGE:3028841 5',
ACCESSION BE299394
VERSION BE299394.1 GI:9183142
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 766)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-romail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM87 row: 0 column: 18
High quality sequence start: 6
High quality sequence stop: 674.
FEATURES
    source
        1. .766
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3028841"
            /tissue_type="rhabdomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 17"
            /note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
            Site 2: XhoI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCAGAG(G). Size-selected
            for average insert size 1.8kb. Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

## ORIGIN

```

Query Match 16.7%; Score 83.8; DB 10; Length 766;
Best Local Similarity 59.4%; Pred. NO. 0.066;
Matches 142; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTGCTCGCTGCACATGCGGGTGCAGGCCCTCGAGCGCGGAGTGCTCC 60
Db 395 TCCTCCAGCCTGATGGCCCTCGAGAACGCTTACAGGCCATCCAGCGGGCAGTGCCCT 336

QY 61 ATGGCGCTCGCGGGCGGCTGATGTGTATGGGACAGTTCGACAGCTTCTGTCAGTTCTCG 120
Db 335 GCCGCCATCGTGGGGGGATCAACGCTCTGTGAGAGCCCAACACCTCCGTGCGAGTTCTTG 276

QY 121 CGGCAACGCGGCTCGCCCCCGACCGCTCGAAGGGGTTGCCGACAGCGCGGACGCG 180
Db 275 AGGCTGGGGATGCTCAGGCCCGAGGCGACCTCGAAGGCTTCGACACAGCGGGGATGGG 216

QY 181 ACCGCTGTGTCGAGGGCGTGGGGTGCTCTCTGTGGAGCGGCTGTCCGAGCTCAGCG 239
```

```

Db 215 TACTGCCCTCGAGGGTGTGGTGGCCCTCTGTGCTGACCAAGAGTCCCTGGCCCGCG 157

RESULT 12
LOCUS BQ366242
DEFINITION BQ366242
ACCESSION BQ366242
VERSION BQ366242.1 GI:21041754
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 275)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Fala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-GN0162-
121000-624-d04&t3=2000-10-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 275.
FEATURES
    source
        1. .275
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="GN0162"
            /note="Organ: placenta normal; Vector: puc18; Site 1:
            SmaI; Site 2: SmaI; A mini-library was made by cloning
            products derived from ORESTES PCR (U.S. Letters Patent
            application No. 196.716 - Ludwig Institute for Cancer
            Research) profiles into the puc 18 vector. Reverse
            transcription of tissue mRNA and cDNA amplification were
            performed under low stringency conditions."
```

## ORIGIN

```

Query Match 16.5%; Score 82.8; DB 13; Length 275;
Best Local Similarity 60.0%; Pred. NO. 0.08;
Matches 138; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 10 CTGCTCGCGCTGCACATGGCGGTGCAGCCCTCGAGCGCGCGAGTGTCTCATGCGCGTC 69
Db 6 CTGATGCGCCTGCAGACGCTTACCAGGCCATCCACAGCGGCGAGTGCCTCGCCGCATC 65

QY 70 GCGGCGCGCTGATGGTATGGGACGCTCGAGAGCTTCGTCGAGTTCTCGGGGAGCGC 129
Db 66 GTGGGGGCGATCAATGTCTCTGTAAGCCCAACACCTCCGTGCACTTCTTGAGGCTGGG 125

QY 130 GGGCTGGCCCCCGCGCGCTGCAAGCGGTTTCGCCGAGCGCGGACGCGACCGCGCTGG 189
Db 126 ATGCTCAGCCCCGAGGGGCACTTCGAAGGCTTCGACACGCGGGGAATGGTACTGCGCG 185
```

QY 190 TCGAGGGCTCGGCTGCTCTGGTGGAGGGCTGTCCGAGGCTCAGCG 239  
 Db 186 TCGAGGGGTGTGGTGGCGCTCTCTGACCAAGAAGTCCCTGGCCCGCG 235

RESULT 13  
 BM091310  
 LOCUS 1921a10.x1 Human Fetal Pancreas 1B mRNA linear EST 20-NOV-2001  
 DEFINITION TR:Q16702 Q16702 FATTY ACID SYNTHASE 1, mRNA sequence.

ACCESSION  
 BM091310  
 VERSION 1 GI:17020276

KEYWORDS  
 EST.

SOURCE  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens

REFERENCE  
 1 (bases 1 to 627)

AUTHORS  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Konko, I., Bennett, J.,  
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R.,  
 Williams, T., Jackson, Y. and Bowers, Y.

TITLE  
 Endocrine Pancreas Consortium

JOURNAL  
 Unpublished (2000)

COMMENT  
 Other ESTs: i921a10.y1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@imgate.wustl.edu)

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 480.

FEATURES  
 Location/Qualifiers

1..627

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/tissue\_type="Fetal Pancreas (4 Pooled Donors, 18 - 20

weeks, Stratagene #738023)"

/dev\_stage="Fetal Pancreas"

/clone\_lib="Human Fetal Pancreas 1B"

/notes="Vector: pBluescript SK(-); Site 1: NotI; Site 2:

XhoI; cDNA made by oligo-dT priming. Size-selected on

agarose gel. Average insert size ~1kb. 5' XhoI site was

destroyed after directional cloning. Amplified once.

Contact information: Hiroshi Inoue, MD, Metabolism Div.

(Alan Permutt Lab), Washington University School of

Medicine, Box 8127, 660 South Euclid Ave, St. Louis, MO

6310 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,

Fax: 314-747-2692."

ORIGIN

Query Match 16.4%; Score 82.2; DB 12; Length 627;

Best Local Similarity 59.0%; Pred. No. 0.1;

Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTCGGCTGCACATGGCGGTGCAGGCCCTCGAGCGCGGAGTGCTCC 60

Db 295 TCCTCCAGCCTGATGGCCCTGCGAAGCCCTACAGGCCCATCCACAGCGGCGAGTGCCT 354

QY 61 ATGGCGCTCCGGCGCGGTGTGGTGTATGGGACGGTTCGAGACGTTCTGTCGATTTCTCG 120

Db 355 GCGGCCATCTGTGGGGGCATCAACGTCCTGCTGAAGGCCCAACACCTCCGTCGACGTTCTTG 414

QY 121 CGGCAGCGCGGGCTGGCCCGCGAGCGCGCTGCAAGCGCTTGCAGCGCGGACGGC 180  
 Db 415 AGGCTGGGATGCTCAGCCCCGAGGGCACCTGCAAGCGCTTCGACACAGCGGGAATGGG 474  
 QY 181 ACCGGCTGGTCCGAGGGCGTGGGCTGCTCTCTGTGTGAGCGGCTGTCCGAGGCTCAGCG 239  
 Db 475 TACTGCCGCTCGGAGGGTGTGGTGGCGCTCTCTGTGTGACCAAGAGTCCCTGGCGCGCG 533

RESULT 14

EX384927

LOCUS

DEFINITION

EX384927 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

Homo sapiens cDNA clone CS0DL004YA22 5-PRIME, mRNA sequence.

ACCESSION

EX384927

VERSION

EX384927.1 GI:30459273

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 988)

AUTHORS

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 8574.r

For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DL004B11Q1&cluster=8574.r. Contact :

Feng liang Email: fliang@lifetech.com/Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID: CS0DL004B11Q1.

Location/Qualifiers

1..988

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DL004YA22"

/cell\_type="B CELLS (RAMOS CELL LINE)"

/clone\_lib="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT

25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 16.4%; Score 82.2; DB 13; Length 988;

Best Local Similarity 59.0%; Pred. No. 0.11;

Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 TCGTCTTCGCTCGGCTGCACATGGCGGTGCAGGCCCTCGAGCGCGGAGTGCTCC 60

Db 652 TCCTCCAGCCTGATGGCCCTGCGAAGCCCTACAGGCCCATCCACAGCGGCGAGTGCCT 711

QY 61 ATGGCGCTCCGGCGCGGTGTGGTGTATGGGACGGTTCGAGACGTTCTGTCGATTTCTCG 120

Db 712 GCGGCCATCTGTGGGGGCATCAATGCTCTGTGTAAGCCCAACACCTCCGTCGATTTCTTG 771

QY 121 CGGCAGCGGGGTGGCCCCCGACGCGCGCTTCGAGGCGTTCCCGACCGCGCGGACGGC 180

Db 772 AGGCTGGGATGTCTAGCCCCCGAGGGCACCTGCAAGCCCTTCACACACGCGGGGAATGGG 831

QY 181 ACCGGCTGTCCGAGGGCTCGGGCTGCTCTGTGTGAGCGGCTGCCAGGCTCAGCG 239

Db 832 TACTGCGCTCGGAGGGTGTGGTGGCCGCTCTGCTGACCAAGATCCCTGGCCCGCG 890

Thu Jun 17 08:25:05 2004

```

RESULT 15
BX332425      1201 bp      mRNA      linear      EST 02-MAY-2003
LOCUS      BX332425 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION      cDNA clone CSODC012YI01 5-PRIME, mRNA sequence.
ACCESSION      BX332425
VERSION      BX332425.1 GI:30335143
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 8574.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC012AE01QP1&cluster=8574.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CSODC012AE01QP1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC012YI01"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 16.4%; Score 82.2; DB 13; Length 1201;
Best Local Similarity 59.0%; Pred.No. 0.11;
Matches 141; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 1 TCCTCTTCGTCGCGCTGCACATGCGGTGCAGCGCCCTGCAGCGCGCGAGTGTCC 60
Db 660 TCCTCCAGCCTGATGGCCCTGCAGAACGCTACAGGCCATCCACAGGGGCACTGCCCT 719
QY 61 ATGGCGCTTCGCGGCGCGGTGATGTTGATGGGCACGGTCGAGACGTTGCTGAGTTCTCG 120
Db 720 GCCCCATCGTGGGGGCATCATGTCTGCTGAAGCCCAACACCTCCGTGAGTTCTTG 779
QY 121 CGGAGCGCGGGCTGGCCCGCGCGCTGCAGAGCGTTGCGCGAGCGCGGACGCGC 180
Db 780 AGGCTGGGGATGCTCAGCCCCGAGGGCACCTGCAAGCCCTTCGACACAGCGGGGAATGGG 839
QY 181 ACCGGCTGTCGAGGGGCGTCGGGCTGCTCTGTTGGAGCGGCTGTCCAGGCTCAGCG 239
Db 840 TACTGCGCTCGAGGGGTGTGTGGCCGCTCTGCTGACCAAGAGTCCCTGGCCGGCG 898

```

Search completed: June 17, 2004, 03:18:50  
Job time : 1966 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:09:53 ; Search time 2276.33 Seconds  
(without alignments)  
9539.403 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggccccagccgagca.....ccctcggcgacagagctga 501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :					GenEmbl:				
1:	gb	ba:	1:	501	100.0	53784	1	AMW223012	AJ223012 Amycolato
2:	gb	htg:	2:	501	100.0	53789	6	A69720	A69720 Sequence 3
3:	gb	in:	3:	501	100.0	90445	1	AF040570	AF040570 Amycolato
4:	gb	om:	4:	275.6	55.0	22131	1	AF506522	AF506522 Streptomy
5:	gb	ov:	5:	268.8	53.7	6567	1	AF521897	AF521897 Streptomy
6:	gb	pat:	6:	164.4	32.8	32870	1	AF007101	AF007101 Streptomy
7:	gb	ph:	7:	158.6	31.7	4689	6	BD217187	BD217187 DNA encod
8:	gb	pl:	8:	158.6	31.7	37948	1	AF079138	AF079138 Streptomy
9:	gb	pr:	9:	158.6	31.7	37948	6	BD217173	BD217173 DNA encod
10:	gb	ro:	10:	158.6	31.7	38506	6	BD232534	BD232534 Recombina
11:	gb	sts:	11:	158.6	31.7	38506	6	AR271680	AR271680 Sequence
12:	gb	sy:	12:	158.6	31.7	38506	6	AR271680	AR271680 Sequence
13:	gb	un:	13:	158.6	31.7	38506	6	AR271680	AR271680 Sequence
14:	gb	vi:	14:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
15:	gb	ba:	15:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
16:	gb	fun:	16:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
17:	gb	hum:	17:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
18:	gb	in:	18:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
19:	gb	mu:	19:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
20:	gb	om:	20:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
21:	gb	or:	21:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
22:	gb	ov:	22:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
23:	gb	pat:	23:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
24:	gb	ph:	24:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
25:	gb	pl:	25:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
26:	gb	pr:	26:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
27:	gb	ro:	27:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
28:	gb	sts:	28:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
29:	gb	sy:	29:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
30:	gb	un:	30:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
31:	gb	vi:	31:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
32:	gb	ba:	32:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
33:	gb	fun:	33:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
34:	gb	hum:	34:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
35:	gb	in:	35:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
36:	gb	mu:	36:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
37:	gb	om:	37:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
38:	gb	or:	38:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
39:	gb	ov:	39:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
40:	gb	pat:	40:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
41:	gb	ph:	41:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
42:	gb	pl:	42:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
43:	gb	pr:	43:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
44:	gb	ro:	44:	158.6	31.1	4725	6	AX697993	AX697993 Sequence
45:	gb	sts:	45:	158.6	31.1	4725	6	AX697993	AX697993 Sequence

ALIGNMENTS

RESULT 1	AMW223012	53784 bp	DNA	linear	BCT 09-FEB-1998
LOCUS	AMW223012	53784 bp	DNA	linear	BCT 09-FEB-1998
DEFINITION	Amycolatopsis mediterranei genes encoding rifamycin polyketide synthases, ORFs 1 to 5.				
ACCESSION	AJ223012				
VERSION	AJ223012.1	GI:2764760			
KEYWORDS	ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.				
SOURCE	Amycolatopsis mediterranei				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.				
REFERENCE	1				
AUTHORS	Schupp, T., Toupet, C., Engel, N. and Goff, S.				
TITLE	Cloning and sequence analysis of the putative rifamycin polyketide				



NYAANAFLDGLMAGRAAGLPVSLAWGLMEQADGLTANLSATDOARMSRGVLPMT  
PAEALDIFDIGLAEQALLVPLKDLRLTGOATAGGEVPHLLRLGLVRASRRVTPTA  
ASGGGLVHKLAGRAPEOEAVILLGLVQAEAAVLPNAPAPAEQATRGFSLDGFSLT  
AVELNRLSAAATGKPLPATLVPDYPTPVALARHLEELGETVAGAPATPVTIVADAGE  
PTAIYVMAARLPVGMVSDPDLRMVWAEGRDMSPPFGDGLDGLDPSDPERPSTAY  
TROGGLFHEAALFDPGFFGISPREALMDPQORLLEASWEALERAGIDPTKARDGAV  
GVFSGVTHDXLESLSNPALEUGFVTTATAGSASGRSVYTFEGEPATVDTACSS  
SLVAHLAAQALROECTMALAGGVAVMSPIGVLMGSRQMGWADGRVKAFAADGAD  
TVLSBGGLVILVERLSVAKERHRLVAVLGRSAVQDGNASGLTAPNGSPQORVIRSA  
LAGAGLQSEVDVFAHGTGTALGEPTEAQAALLATYKGRSTPLMGLSLKSNIGHTQA  
LAVAGLQSEVDVFAHGTGTALGEPTEAQAALLATYKGRSTPLMGLSLKSNIGHTQA  
AAGAAVAKVMAKQDPTLPTTHQEPTEQVDSAGAVELLTSGREWARNGHPRRA  
VSGFTSGTNHLLLEAEADPTADVPDVPVVISARSTGLAQAGRLAFLDG  
DVELTRVAGALLSTRATLTVRAVAVVPSAEERAGLTALARGESASGLVTGAPMGK  
TVWFPQGTQWAGGRELLESPVFAERIEECAAALOPWIDLSLLDVTGEGELDRV  
QIAAALSGRGWASIQSHDEVAALQAPWAGRVETAAVNGPASPVIAGAEALTEAV  
EVLGRRVADVSHTRVEDIQTDLAETLAGIDQAQVVPVFTSVIAGEWITDAGVVD  
GGVWRNLNQVGFVPAVVAELIEQHGCVFVEVSAHPVLQPISELTDVAVTGLRRDD  
GGVRLRLTSMALFVRGVPDMATAPPAVELPTVAFDQHFWLSPPPVADAPALGL  
AGADHLLCANLPLPQSDGLVFTSRLSVRTHPLADGVPAALVELAVRAGDEAGCPV  
LADLVEKULVLFPESGGLRVQIVSGERTVYVYVQLEGAEDWIRNATGHLSATAPAH  
AFDFTAPPAQAQVQDGLMRRGEIIFAEVALPEELDAGAFGIHFPLLDRAVQVPLADD  
KQPAEWSLVLHRAAGSALRLVPGGALQAADETCGLVLTADSVAGRELISAGKTRAG  
SLVRDWTESVTSADSAVPAVIEVEAFGEPELTLGRVLEAVQTLVPTAADARLVV  
TSGAVREVPDPAAGVWGLVRAQAEENPGRIPLIDTDGEIPALTDEPELTAVRGGKFF  
VPRITRABSPGAAPVPRDGTVLISGAGLGLVARRLVERHGVKLVASRRGRDDVG  
VADLADLAADSVVACVSDRAQVAAALLDHRPTAVVTHAGVIDAGVLETLDRDLA  
TVFAPKVDVAVHDELTRDLDLDAFVYSSVAVFVAGSGVYAAANAFDGLMANRR  
AAGLGLSLANGLWDSQSTMAAGTDEATRAMSRGGLQIMTQAEGLDPLDALSSAR  
SLVPLKDLIRAGDAAGVPHMLRGLVRAGRAQAAASTVDRGLRGLAGLAPA  
DOJTLALLDIAVRAVAVLGHADASVRVDTAFKADGDSLTAVELNRMTATGLKLP  
ATLVPYPPQALABHRLDEGNAOTPVTTAAKADLDEPIALVGMACRLPGVAGP  
EDLWRILVABGRDVAESPTRDNDTSLYDDPARPKTYTRHGGFLHEAGLDAGFF  
GISPRRAVMDQORLLEASWEAMEDAGVDPLSKNDVGVTFGRFGQCYVASRQSV  
VTEPEFAGTGGSSVAVSRVYVFGFEPVAVTIDSSSSLVAMHLEAQAIRQREB  
SMALAGATVMAVPGAFVRFSGRLAVDRCKAFVAAADGTGMAEGLVAVVILLERUSV  
ARRERHRI LVAIPGSAVNOAGNGLTAPNGSPQORVIRREALVSGALPDDVDEAH  
GTCTTGLDPTLEAALLATYKDRSPLMLGSLKSNIGHAQAAGVAGVIMVQALRIE  
VLPPTLHVDTEVDWSAGAVELLTEAREPNRPNRPRAGVSPVSGVGNHILIE  
APAEFPVPTPEVPLPVVVSARSASLAGAQRGLAGFVAGDASLAGVARALVITNRAAL  
TERAVMVVGRREEAVTNLEALARGEDPAAVVTGRAGSPGKLVVWVFPQGSQWIGMRE

Query Match 100.0%; Score 501; DB 1; Length 53784;  
Best Local Similarity 100.0%; Pred. No. 1.4e-43;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCGGTCCCGACGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCAACGCTTCG 60  
Db 15043 GCGGTCCCGACGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCAACGCTTCG 15102  
Qy 61 ACGGTGCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 15103 ACGGTGCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15162  
Qy 121 GGCTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCGCG 180  
Db 15163 GGCTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 15222  
Qy 181 CGGCTGCGCGCGACGCTGATCTTCGACTACCCCGCGGAGGCGCTGGTGGCTTACCTG 240  
Db 15223 CGGCTGCGCGCGACGCTGATCTTCGACTACCCCGCGGAGGCGCTGGTGGCTTACCTG 15282  
Qy 241 CGGCTGCGACTCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 15283 CGGCTGCGACTCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15342  
Qy 301 CGAGTCTCTCGCGCGCGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 360  
Db 15343 CGAGTCTCTCGCGCGCGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 15402  
Qy 361 CTGCGCTCTCGCGCGCGACCGAGCGCTGATCGACGCACTGCGACATCTCCGCTTCCTGCAACGA 480  
Db 15403 CTGCGCTCTCGCGCGCGACCGAGCGCTGATCGACGCACTGCGACATCTCCGCTTCCTGCAACGA 15522  
Qy 481 GCGCTCGGCGCGACGAGCTGA 501  
Db 15523 GCGCTCGGCGCGACGAGCTGA 15543

421 CCGCGCGCGCGACGACGCGAACTGATGACGCACTCTCGGCTCTCGTGCACGA 480  
15463 CCGCGCGCGCGACGACGCGAACTGATGACGCACTCTCGGCTCTCGTGCACGA 15522  
481 GCGCTCGGCGCGACGAGCTGA 501  
15523 GCGCTCGGCGCGACGAGCTGA 15543  
RESULT 2  
A69720 A69720 53789 bp DNA linear PAT 07-MAY-1999  
LOCUS Sequence 3 from Patent WO9807868.  
DEFINITION A69720  
ACCESSION A69720  
VERSION A69720.1 GI:4774328  
KEYWORDS . unidentifed  
SOURCE unidentifed  
ORGANISM unidentifed  
REFERENCE 1 (bases 1 to 53789)  
AUTHORS Schupp,T., Toupet,C. and Engel,N.  
TITLE RIFAMYCIN BIOSYNTHESIS GENE CLUSTER  
JOURNAL Patent: WO 9807868-A 3 26-FEB-1998;  
CIBA GEIGY AG (CH)  
FEATURES Location/Qualifiers  
source 1..53789  
/organism="unidentifed"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 100.0%; Score 501; DB 6; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 1.4e-43;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCGACGCTTCG 60  
Db 15043 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGCTGAAGCTGGTCCGCGGCGACGCTTCG 15102  
Qy 61 ACGGTGCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 15103 ACGGTGCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15162  
Qy 121 GGCTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 180  
Db 15163 GGCTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCG 15222  
Qy 181 CGGCTGCGCGCGACGCTGATCTTCGACTACCCCGCGGAGGCGCTGGTGGCTTACCTG 240  
Db 15223 CGGCTGCGCGCGACGCTGATCTTCGACTACCCCGCGGAGGCGCTGGTGGCTTACCTG 15282  
Qy 241 CGGCTGCGACTCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 15283 CGGCTGCGACTCTCGGCGCGCGAGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15342  
Qy 301 CGAGTCTCTCGCGCGCGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 360  
Db 15343 CGAGTCTCTCGCGCGCGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 15402  
Qy 361 CTGCGCTCTCGCGCGCGACCGAGCGCTGATCGACGCACTGCGACATCTCCGCTTCCTGCAACGA 480  
Db 15463 CCGCGCGCGCGACGACGCGAACTGATCGACGCACTGCGACATCTCCGCTTCCTGCAACGA 15522  
Qy 481 GCGCTCGGCGCGACGAGCTGA 501  
Db 15523 GCGCTCGGCGCGACGAGCTGA 15543

```

RESULT 3
AF040570      90445 bp      DNA      linear      BCT 16-APR-2001
LOCUS      Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
DEFINITION
ACCESSION      AF040570 AF040571
VERSION      AF040570.2 GI:123331604
KEYWORDS
SOURCE
ORGANISM      Amycolatopsis mediterranei
              Amycolatopsis mediterranei
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Pseudonocardineae; Pseudonocardiaaceae; Amycolatopsis.
              1 (bases 1 to 90445)
              August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,
              Taylor,M., Hoffmann,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and
              Floss,H.G.
              Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
              the molecular analysis of the rif biosynthetic gene cluster of
              Amycolatopsis mediterranei S699
              Chem. Biol. 5 (2), 69-79 (1998)
              9512878
              2 (bases 1 to 90445)
              Kim,C.G., Yu,T.W., Pryhle,C.B., Handa,S. and Floss,H.G.
              3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
              formation of the precursor of mC7N units in rifamycin and related
              antibiotics
              J. Biol. Chem. 273 (11), 6030-6040 (1998)
              98165773
              3 (bases 1 to 90445)
              Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,
              Leitner,E., and Floss,H.G.
              Mutational analysis and reconstituted expression of the
              biosynthetic genes involved in the formation of
              3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
              biosynthesis in amycolatopsis mediterranei S699
              J. Biol. Chem. 276 (16), 12546-12555 (2001)
              21201076
              4 (bases 1 to 90445)
              August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
              Hutchinson,C.R. and Floss,H.G.
              Direct Submission
              Submitted (30-DEC-1997) Dept. of Chemistry, University of
              Washington, Box 351700, Seattle, WA 98195-1700, USA
              5 (bases 1 to 90445)
              August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
              Hutchinson,C.R. and Floss,H.G.
              Direct Submission
              Submitted (19-JAN-2001) Dept. of Chemistry, University of
              Washington, Box 351700, Seattle, WA 98195-1700, USA
              Sequence update by submitter
              On or before Jan 22, 2001 this sequence version replaced
              gi:2792340, gi:2832263.
              Location/Qualifiers
              1. 90445
              /organism="Amycolatopsis mediterranei"
              /mol_type="genomic DNA"
              /strain="S699"
              /db_xref="taxon:33910"
              423..1451
              /codon_start=1
              /transl_table=11
              /product="unknown"
              /protein_id="AAC01706.1"
              /db_xref="GI:2792310"
              /translation="MAPTGGGSEYRYKTLQLEVALIGALILHLEAWLGAAGV
              RYSDGSRLEAEFGAKAGSLEALDGDADVDICTPTASHHEIALTATAAGGVV
              CEKPLAAAEAEIEIVTAERAGVRLYAAHDVHFAAPYARLHELHVSRLGEGALGRF
              SFSYHPHPWTCHASASGGILITDQLLGHADLAHWFDGVVRVHACYGDILATPAEP
              AVATCTALTHASGAI SOWSRKWATATPPPVRFVPHVSGTGSYSYDSEWQEVVRVD
              GGAGNFAGGVSVPTEMRERFATAPAGGPEPRIGAKDALAAVRIHAAEASAWTGRAV
              ELFVRGAA"
              CDS
              1448..2215
              /codon_start=1
              /transl_table=11
              /product="unknown"
              /protein_id="AAC01707.1"
              /db_xref="GI:2792311"
              /translation="MKVAILSSSTPQAYAGLRGLPVEVVAASWDAFEPVRAEAG
              ARVLCYPPPAKETDLKAMIDAGRLTFASPACHGAFAPVVRKIGADGGELITTVL
              GSVATSDVGLGAAAYLLDADAVLAGEPAQVYQNTNIVLSRIGESAALIVRYR
              SGVASDFDCRRHGSATGLPAVTFLGDQSVQYDAGPQLGGERPELGEDLEALMLKD
              FLGADGPGPGQALRTFRILQIAYESAHTGQFVDL"
              2353..2619
              /codon_start=1
              /transl_table=11
              /product="unknown"
              /protein_id="AAC01708.1"
              /db_xref="GI:2792312"
              /translation="MQDRSVLGDPPKQVEAAMSVTQVKIALSGGPAELAPAHRSVDSG
              ALENTKIRHGAGYEHFVHGGEFOQAVDGCDAVFRWSHRTKIAE"
              2652..3842
              /note="ORF0; similar to Streptomyces Sp. Chop"
              /codon_start=1
              /transl_table=11
              /product="cytochrome P450 monooxygenase"
              /protein_id="AAC01709.1"
              /db_xref="GI:2792313"
              /translation="MTDAISFEVPWMDRTDKFPPAVDFSLREERFLAKMVPDGHVGM
              TVSSYELREVLSDLRFSGHSCVGHFVTHQQVTPHPLIPGMTHMDPPHTRYRK
              LITGFTVRRASRLIPRAEVAAGCIEVMKAGADVMDFAKPLVLMGLSELVGLP
              YEERKYVPAVTLHDDADPAEAAAYVAGKFFDEVTEERRQPDQLISLVTED
              LTQEELRNIVTLFAGYETTGALATGVFALLHTDQLAARAEPEKLDAALELLR
              VLVNQYHTYRTALEDKLEGLIKKGDVTVLSLPAANDPKAFKFCFALDTERDTS
              HVAFGFIHQCLGQNLRIELRAGETALLRAFPFELRLAVPADEVLPRLKGVSPVSKKL
              PYSW"
              3883..18090
              /genes="rifa"
              3883..18090
              /genes="rifa"
              /note="Rifa"
              /codon_start=1
              /transl_table=11
              /product="polyketide synthase"
              /protein_id="AAC01710.1"
              /db_xref="GI:2792314"
              /translation="MRTDLIKPLHVALLENATFACKGAFADHRTVTYVGGDLERTRR
              LAGHLAGVRRHGRDVAICLGNRVSVESVFAIRAGAVGVLNPGSATAAEHLPTD
              SGATVVTDAQAARLRLAPHVELLVTDGDPPEGAHSYDELAISEPAADLELDE
              PAWFTYSGTIGPKGVSTQRCNLMSVASYVPFPLSDQDRVLPPLPLFSLSHIA
              CVLSATVGASVRIADSSADDVRLIEASSTFLAGVPTTVHILVRAARQGFSAFS
              LRIGLAGAVLGAHRSFEETFGVLIDRAGSTETGCGALTMPPGARVSGGLAV
              PGVDRVDPDTGLDVPAGEEVGPNVLMGYHNSPEATAAMRDGMFRFTGDLAR
              RDDAGYFTTCIKELIIRGGANIHPEVEAVLTVGDVADAAGVGVGPHDTLGEVPA
              RYTPGPTDPAALIEKREOLSAKYVPDRILEVAHIPRTASGKIRRLGLTDEPAQLR
              YATHEEQSRHADESVALRALRALSGLDERAOCELLEDLVRTQAADVLGPQPDGRA
              FRDLGFTSLAIVELNRNLTEHTGMLPASVDFHPTPALAARVRAELIGITQAVAP
              VVAADPEPTAIYVMACRLPGGVASPEDLMLVAERVDVDSFEFPGRCMDLSDLPD
              RERAGTSYQGGFLDADGEFAGFFGISEAVAMDPOQLLELTSWEALNKGVD
              IALKGTDTGVFGLMGQYGSAGVAPLEGVTTGVASSVSGRSVSLVLEGPAVTV
              TACSSLVAMHLAAQALROGCSMALAGGVTVMATPGSFVFSRQDRGCKAF
              RIVRRVAVDVAISHRTHVEDIDTLAETLAGISAQAPVPTSTVTSWDRDAGVLDG
              YWYRNLRNQRVFGAAATALLDQGHTEVFEVSEHPVTQPLSELGTDAIGTLRRDGG
              RLILASMGELFVRGIDVDTWTMPAAGWDLPTVAFEHRYHLEPPEPAGAGVPLIGT
              VVLTSPGDRLLTVAQWRRACQPMVADGLVNPAAALVEAAIRGLDAGTVPVGLVDFAP
              VLTPRRSGREVQLIIGEGGQRRRPLEVFSEADPEWTRHAGTGLAATAAAVPPPA
              GDATDVTVAGLRDADRYGIHPALLDAAVTVVGGDILLPSVWTVGSLASGATATVTT

```



TATGRLRLTDPAGQVLTIVSVYRGTGPPVAEGCTTDTALFRVDMPEIPLPETAETADFLPYE  
 ATSAEATLUSALQAWLADPAETRLAVVVGDCTEPGRALWGLVRSQSEHPGRILVLADL  
 DDPVLPVAVASQVQVVRNGVASVPRITVPRQDARPLDPGTVLITGCTGLGA  
 LTARHLVTAHGVHVLVLRGEAPLELTAELTALASVAIAACVADRAEVLAVLAI  
 PAEHPETIAITAGVLDGVVTVLPEDRLATVRRPKVDAAARLLDBLTREADLAAFLV  
 SSAACVGNPGQAGYAANAELDALARNOLDLPAVSLAWGYMAVTSGMTBHLGDAL  
 LERNORI GMSGLPADEGMALLDAAATATGCTLVAEKEDVAALPATAKAGCPVPLLRGL  
 APLPRAAATASLTSLAGLAETGQAALLDLVRRHAAEVLVCHSGASVHSGRIFPKD  
 AGFSLTAVELNRLAAATGLTSLPAMIYDYPKPPALADHLRAKULFGSAAANPABIGT  
 AAAEPIAIAMACFPQGVHSPEDLWRLVADGADAVTEFPADRWGMDTLRYHEDPDH  
 AGTTLVGFAGVNSHDAGDAFAFGISPEALAMDPOQLLLETSELWELFERAAIDPTTL  
 ECODLGVFAGVNSHDAGVSMHRAAGVGFRLTGGASVLSGRVAVHFGVEGPAVTVDT  
 ACSSILVALHMAVQALORCECSMALAGVMMVMTVETVFERFQRGLAPDGRCKAFAD  
 GADGTGNSGVGLLILVERLSEARQHQVLAIVRGSAVNSDGSNGLTAPNGPSQOV  
 TRKALAAAGLSLVDVAVHAGTGTTLGDIIEAEALLATGQNRTEPLWLSVKNLIG  
 HTCAAGVAGVITKMMHMHVLPRTLVDRPSSVYVDSAGAVELLTEARDVNSNPH  
 RAAGVSSFGITGNVHVEEVAAPITTPQBPFAELVFLVLSATAAGLQAGRLA  
 AFLGRTDVRVPDAAVALATTRAQDHRVAVLASDRAQLCADLAAFGVGTGTPVDG  
 KLAFLPTGQSGWAGNRELATFPVFRDAFPAACVAVDTHLRERFLRWVFDSDALL  
 DQMTYTOGALFAVETALFRLFESWVRPGLLAGHSIGELAAHVSGLVLDADAGELVA  
 ARGRLMQALPAGAMAVAVQATEDEAPLDDGTVCVAVNGPDSVLSGTEAAVLAIVAD  
 ELAERGKTRRLASVHAFSPLEMPMLDDFRAVRELTYRAGSLPVSTLITGELAALD  
 SPDYVWGVNRAVFDATAVTAQAGASTFLELPGGALAAMALGTGEPQSCVATL  
 RKNRGAEPDVLTALEHVRGVGDMTVLDEPATVAVTVLPYAFQHORFVVDVDET  
 KURNRLDTGTVELPSTLIFDYPNPAVLAHLRAELLGERPAPAPVTRDVSDEPIAI  
 VGNSTLPGQASPESELWKLVAEGRDVSFVDRWGLDDGLTHPDPAHAGTSYTRSG  
 GFLHDAQPDAGLFGISPREALAMDPOQLLLETSELWELFERAAIDPTTL  
 GIVHNDYVTRLEVPDQVYTMGTASSVAGVAVYFVGEPVAVTVTACSSSLVA  
 MHLAALQAGCECMALAGATVMAWSPDAFLFESRQGLSADGRCKAYAEAGDGTGWA  
 EGVGVVLRFLSVARERHVLAVLRSVAVNODASNGLTAPNGPSQOVIRGALASA  
 GLAPSDVVEHGTGTALGDIPIVQALLATYQEREQPLWLSLKNLGHGTCAAAGV  
 VGVIMAMRHGMVATILHDEKTSQVDSAGAEVITHEARWPRTRGRPRRAGVSF  
 GASGTNAHLIEEGPAAEVEDEVSVPLVSVARSAGLAGQAQALVAENESLAG  
 DGVRVAVAYASHYTHVBAIATLAKTLADARVAPLFPVSTVLTGWLQVNDVA  
 GYNNRLQOVRGSPVADVAGLGHVTVFVISAHPVLVQLSEISDVAATVSGSRRDD  
 GGLRLILAAELVAVVAVDVAAPVAGVLDLTPYAFDRHFWLHEARSTAEKRLQVY  
 DGEFTWAIKOSDLSVLAELVPEQRLSTVVPVLAQWRDRRRSTAEKRLQVY  
 WQLERAAGVPGGRVAVVAPAGTITDALLKELTQGLDVLRLTEASRAQLAOLRN  
 VLEHDTGLSLALDGGPADAAETASTIALVQALDQTTTSAPLWCLTSVAVNIGI  
 QDVAIPAQAAGVGLKRAVALERDRWGLVDPALADARTQAQLGVNKAAGEDQL  
 AVRRSGYRRRLVRKVPESATSWEPRTVLTGGAEGIGRSHASVWLAGSAGRLIV  
 TGTGDVDELTAELAEFGTVEFCADTDRDAIQLVADSEVTVAVHAADIATGSVDVT  
 GVADLDFVAAKTVAVLQDLPEDTDLPAVFPVSSVAGIAGVGGGQGPAGAANAVLDA  
 LVWERRARGLKATSIANGALDOIGIGMDEAALQOLRRRGVTPMAPPLAVTAMVOAVAG  
 NEKAVAVADMWAAFIAPFVSVPSPPLPADLPEAKAILRAAQDDGEDGTASSLADSL  
 RAVPDQNRLLKLVHGHASTVLSHSGARGIGRQAPQEVGFDSLAANVLRNLSLHAA  
 TGLRLPATLIFDYPTEPALVGYLRELLREADDGLDGRDRLVRLVAVPAPFKEAG  
 VLDTLGLADTGTGTEPDAETTEAAPAAADDAELIDALDISGLVQALGQTS"  
 3883..5775  
 /gene="rifa"  
 /note="encodes polyketide synthase loading domain"

Query Match 100.0%; Score 501; DB 1; Length 90445;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-43;  
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGGTCCCGAGCGCGAGCAGAACCGCATCTCTGTAAGCTGTCGCGGCACGCTTCG 60  
 DB 17590 CGGGTCCCGAGCGCGAGCAGAACCGCATCTCTGTAAGCTGTCGCGGCACGCTTCG 17649  
 QY 61 ACGGTGCTCGGCCACAGCGCGCGCCGAGACATCGCGCCGCGCGCTTCAGGAGGTC 120  
 DB 17650 ACGGTGCTCGGCCACAGCGCGCGCCGAGACATCGCGCCGCGCGCTTCAGGAGGTC 17709  
 QY 121 GCGTTTCGACTCGCTCGCGCGCGCTCAACCTCCGCAACAGCTGTCACGCGGCCACCGGGCTG 180  
 DB 17710 GCGTTTCGACTCGCTCGCGCGCGCTCAACCTCCGCAACAGCTGTCACGCGGCCACCGGGCTG 17769

QY 181 CGGCTGCCCGGACGCTGATCTTGAACACCCACCCGAGCGCTGCTCGGCTACCTG 240  
 DB 17770 CGGCTGCCCGGACGCTGATCTTGAACACCCACCCGAGCGCTGCTCGGCTACCTG 17829  
 QY 241 CGCGTCGAATCTCTCGGGAGCGCCGACGACGCGCTGAGCGGGGAGACGACCTCCGG 300  
 DB 17830 CGCGTCGAATCTCTCGGGAGCGCCGACGACGCGCTGAGCGGGGAGACGACCTCCGG 17889  
 QY 301 CGAGTCCTCGCGCGCGCTGCTCGCGCTTCAAGGAGCGCGCGCTGCTGACACGCTG 360  
 DB 17890 CGAGTCCTCGCGCGCGCTGCTCGCGCTTCAAGGAGCGCGCGCTGCTGACACGCTG 17949  
 QY 361 CTCGCGCTCGCGCACACCGCGCACCGGACCGGACCGGACCGGACCGGACCGGCGCC 420  
 DB 17950 CTCGCGCTCGCGCACACCGCGCACCGGACCGGACCGGACCGGACCGGACCGGCGCC 18009  
 QY 421 CGGCGCGCGCACGACCGCGCAACTGATCGACGCACTCTCCGCTCTCGTGCACGCA 480  
 DB 18010 CGGCGCGCGCACGACCGCGCAACTGATCGACGCACTCTCCGCTCTCGTGCACGCA 18069  
 QY 481 GCCTCGCGCACGACGCTGA 501  
 DB 18070 GCCTCGCGCACGACGCTGA 18090

RESULT 4  
 AF506522 22131 bp DNA linear BCT 30-MAY-2003  
 LOCUS Streptomyces hygroscopicus AHBA biosynthetic gene cluster, partial  
 DEFINITION sequence.  
 ACCESSION AF506522 AF506523  
 VERSION AF506522.2 GI:31211262  
 KEYWORDS Streptomyces hygroscopicus  
 SOURCE Streptomyces hygroscopicus  
 ORGANISM Streptomyces hygroscopicus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 22131)  
 Gao, Q. and Wang, Y.  
 Identification and analysis of two separate AHBA biosynthetic gene  
 clusters in Streptomyces hygroscopicus 17997  
 Unpublished  
 REFERENCE 2 (bases 1 to 22131)  
 Wang, Y. and Gao, Q.  
 Cloning of an aminocyclitol biosynthetic genes from  
 Streptomyces hygroscopicus  
 Unpublished  
 REFERENCE 3 (bases 1 to 22131)  
 Wang, Y. and Gao, Q.  
 Direct Submission  
 TITLE Submitted (26-APR-2002) Microbial Pathway Engineering, Medicinal  
 Biotechnology, Tiantan, Beijing 100050, China  
 JOURNAL  
 REFERENCE 4 (bases 1 to 22131)  
 Wang, Y. and Gao, Q.  
 Direct Submission  
 TITLE Submitted (10-DEC-2002) Microbial Pathway Engineering, Medicinal  
 Biotechnology, Tiantan, Beijing 100050, China  
 JOURNAL  
 REMARK Sequence update by submitter  
 COMMENT On or before May 30, 2003 this sequence version replaced  
 GI:24637555, GI:24637552.  
 FEATURES  
 source  
 1. 22131  
 /organism="Streptomyces hygroscopicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1912"  
 <1..2412  
 /codon\_start=1  
 /transl\_table=11  
 /product="Shnc"  
 /protein\_id="AAP45191.1"  
 /db\_xref="GI:31211263"  
 /translation="XDEEDNVRLLTVVSTLRDEGGPRLLTSMALFVRGVFVDW  
 TKFLMAGHAGVLDLPTVAFDRRHYLQDAATADDSGASNDADADFWSAVEQTADSLA

CDS



GLLAPDSAGLDELRTVVPALADWRGSRSSAERLRYATWTPPLDREYSVPAGEW  
QAPVPPCPAETDGTGSRVABLIABLAGAGLDVVPETVPNSAFRTTGLTARLSDIRAEY  
OPAGVLSVLLAQDQDDVIDTARTLALQALGDAGVNGPLMCLTRGAVNTGIGQITAGE  
PDAALWGLGRAALHPDRMGWGLDIPATADHTAQYLVALNGTAGDQLAVRPGFL  
YSRLVRKPSQTPADGSMRPHGVTVLVGGAEALGIIHASLWARGARLITVITTAQ  
PADAVTEQGLKAAAGVETTVVSCADARETLARLIATPREOPLTAVHAADAPWTS  
AVADTGHLDTEVAGKVDPAWMLDELFTGTDAAPLDAFVVFSIAGIWGAGGOGVSG  
AAGVILDALVDRRGRGLAATSIAWGLDGLGMDZAAAQALRRRGLPMHQAQVAT  
AFEGAAREKAVTVAQDMWEAFIPAFTSARVSEPLFADLPAAALRSQPDABNGDI  
TSSVLSDRVPQAEQNRLLRLVGGQAATVILGSSGESIGLPGFOEVEGDSIGAVN  
LRSNLHVAITGLPATLVFPTDPAVVGFLRSLLTETSDDEGREDLRLVLAQVP  
LSRLREAGLLDTLLSLGDSVDSPEAAAPPAAPAAEDAAVIDVMDVADLVKRAL  
GSNNP"

2461..8979

CDS

/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45192.1"  
/db\_xref="GI:31211264"

/translation="WAPDEQIVDALRASLKENMELQENORLSESSAEPIATVSMAC  
RYAGIRNPEDLRVNDNDVYTSFPENRGWDLGIIHPDPNPGTIVREGAFIHD  
ANLFDAGLFGISPLEALAMEPQORLLEICWEALERAGIDHDSVRGADIGVYALHQ  
DYAPDLGLEGLTIERALGSGAGIASRVAYTLLEGPAVTVDMCSSTLVAHVAT  
QALRGECAALAGATVMTSTPGGFIGPARQALAFDGRCKSYGAADDSWAGAGV  
VILLESARNGHEVLAVIRGSALNDQASNGLTAPNGPAQRVIRKALENAGLTVA  
DIDVEHGHTGTVLDPIEAQALJATYQDRPEGPPLMLGSVKVIGHTSGSGVAGL  
INAVQALRHGVPATRHVDANPQVDNAGAGVELLTERANPELORPRRAGVSFSGAS  
GTNAHMLLEQAPPEAPASPSAPADGVVPIVLSAATASLTGQAEIRGSLFEASGTV  
ALADVAALITGRASLAQRAVVVDSPEEALAGLALGARDGVRGAVAGGVRSOGDG  
KVVLVPPGQSPVMWGMGRLELCESEVFAARVECAVALERWMDSLVDVLRGDCPVEF  
FEREDVROPASFAWVGAAWVESVGVVADA VVHSGGEVAAACVSGALSLEDAVRV  
AVRSKISGVLSSGRGMA SVGISSEEAVALQOMDGRVEI GAVNPSPSVITADTEAL  
DPAIETEDQGVVRRIADLYASHSRHGVAGVEILNEAFDIRSQAPVTFSTATGE  
WIREAGALDQVYRNLRKSQVRPAIDLLADGHTVFESSAHFVLPVQPISEVAVGA  
EAEAVTGSRLRHGGPRGLFTSMADLFVYRGTHVDSVGLAACGADARVDVLTYPDH  
KNYMWELAGTANDVASLIGSGADHLLGAVVPEVPTSGVLTSLSLRTHPLWADHAV  
RSVAVYSARDDGRTGIDGWTTHAXGTLTAAVPAKXDFFTVWPVGAERVKEDAAE  
FYEDLADRGYTVGPAGLGRVWRGEEVFVAVSVLPDEORGEASGYGLHPALLDALH  
SGFQVPAGEAIGRVFPAWNLGHAAGSELVRVTFVSGPDAVLSQADAGAGGVLS  
LQSLISRPVQQLDMATDGRDALFEVDMTLPAPAGAPLPFSWTLVASTEDVTE  
LAGATEVPAAVLAQALGGDGHADALITVRLEIQAHPAAADLAESRLVLTQGAVP  
ADGEDVSDPAGAAWGLVRAQAENPDRI VLLDIAANSGLDLGAGMGSVLTALAA  
DPQVAVRTTISVRLPAGKRALXIPEAKXEWRLGVTVGTGLDLSLXPEXXXA  
PLAGQVRVLRGALGNFRDVLILGMYFOAAKXGEGAGVLEVEGVTGLAVGDEV  
MGLTIGYGSMTVDHRCVLRIPQGSVFOAAETVPVFLATYIGLRLDGLAGRAGESVL  
THAAGVGMAATOLARHIGARVFTGASPGKWGLVLELFDXAGTASSATSEFEBHL  
TGTSGAGMDVLDCLAGELVDASRLLPGRGFIEMGKTDIRDAGVASEHFGVAYQA  
FQVFDAGLDRIOQMLLELARLFETGDLTLPFAAMDVRAPADFRHMSQARHTKMVL  
SVYPEPLDGTVTLTGTSGLGAVARHLVGEHGVHVLASRRGPDASGARLVAEL  
GEAGATVAVACDVTRBAKXALLAAXPAEHPLTGVVHXAGVLDGVIOTLTPERVA  
VFAPKVTAVRHLDLHDMDSNFVSSAAGVFGAGQNGYGAANAFLDGMWARRRA  
EGLPGITAMAGLMEQTGMTANLSQADQARMGRGGLPIGVAEGMELFDTALRMPAL  
AVPIKDLRGMRAADAGGSVLALLRALVQTRQQAAGENGSLGARLASLTPOQE  
ELLIDLVTQAATLVGTETISNVAHKA FNAGFDLSITSVELNRLREATGLKLAATS  
VEDYPTPLALARHLHDLKPAQSPCTHPLLAELSLEARLNETPVDSDSIRAQVTRL  
QVLLTWTSTANGPTGDEALDFDATSDDLLFDLIDTFEN"

9111..13136

CDS

/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45193.1"  
/db\_xref="GI:31211265"

/translation="MLELEAKSQPVAIVGMA CRFPDGVASPODLWRLVSEGRGLSG  
FPFGCRWDEFLDAPFRAGKSYTDRGGFHAGLFDAGFFGISPREALAMPQORL  
LLEVSWEALERAGIDPGLTGSVDGVSGVITQCYGMRGETPTELGLXGTGXAXSV  
SGRVSVYFGPGSVSDTACSSLVAMHLAXQALROCESMALAGATVWNPXTFV  
EFSRXXXSDGRCXASAYADCTWGEGAGVVLLELSEARNHQVLAVIRSSAVN  
QDAGNSGLTAPNGFSQVRVIRQALASGLSAADVVEGHTGTGVLGDPTEAQALLAT  
YQGEREPERFLWLGSSNLIGHTQAAGAVAGVITKWVLEALRGVWPATLHADTFSSQVD  
WSAGAVELLTEARDWPTDTPRRRAVSFSMSGTNAHLILEHSEPEQASVAFVSDV  
VPIVVSARSARSLVGQORLASFVGSASDVALPEVAGLVSRRAVLSERAVVAVGSRD  
QALLSGLALARGESHFGVTVTGGSLGGVSGRKVLVPPGQSPVMWGMRELLECSAVFA

ARVGECAVELERWDMSLVDVLRGDCSVFFEREDVRQPASFAVMVGLAAWVESGVV  
ADAVHSGSGEVAACVSGALSLEDAVRVAVRSQIISGVLSSRGSMASVGLSEEAHV  
ARLQOMDGRVEIGAVNPSPSVITSGDTEALGOVLDALAAAGVARRMEVDPAHSROV  
EAEELIRAPADIRAPALPIPFYTSVTGWEVAGVLDGQYVWNLRGVFCGPAIA  
DLADGHTLVFESSASPALVWLINETADENANTEVLVSGSRFGGGLHRLFTSWADLF  
VVRVDMVSGVLAAGADARRVDLPITYAFEFHEHWWLPTMTSTDAASLGADGCPHLLGA  
WVAVPTGCVGLTGSLSLRLTHPLMADHAGVGLVLPCTGLXELVVRADGDEVCCTLEE  
LITBXPVLVPAQGXVOKXAVGGPEENGXRSVAYSARDDBGXTGDXGDMWTRHAGTL  
TAAEPADGDFVTWVPVGAERVSFDVAFVGFVEMAGRGYVYGPAGLGRVWRGEEV  
FAEVALDEQGEASRGLHPALLDALQSLAVPADAVDMRVPEAWNGLRLHAAGA  
SELNRVTPSGSDAVSLQALDAGCPVLSLESLVARAVDVLEQDMATDGRDALFEV  
DMSLTPASPSSVESLPPSALVASAEDVTDLADAAVPAVAVLEAVGCGDEHDAALTV  
RVLEVQAFAAAGLAESRLVVTTRGAVPVGEGNVDPAAGAAWGLVRAQAENPDR  
IVLLDLAADVDMGSLVLPSCXG"

CDS

13847..14533  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45194.1"  
/db\_xref="GI:31211266"

/translation="MPLSPAKESKMTAEKLLKSLKLVTAELHDLROQVARRTDEPV  
ALVGMACRLPGCVSPQDLWRLVSEGREGVSGPEGRGMDVEGLFDADPGRAGKSYTD  
RGGLHAGLFDAGFDGFGISPREALAMPQHXKLARACGCGSTGFCGLPRWRCSWTCG  
GCVLQRPAGVPFPCCVLWCLIPVGRRRRPGTAVELPRGSPVSHRRSRKSCCSTWSGH  
RWRSSSATSR"

CDS

14963..15505  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAP45195.1"  
/db\_xref="GI:31211267"

/translation="MPLSPAKESKMTAEKLLKSLKLVTAELHDLROQVARRTDEPV  
ALVGMACRLPGCVSPQDLWRLVSEGREGVSGFPEGRGMDVEGLFDADPGRAGKSYTD  
RGGLHAGLFDAGFDGFGISPREALAMPQQAAYWEMCTHPASRRSRSDRASSRSVSL  
VSPITSGPPAVVSTPMAPR"

CDS

15358..16812  
/note="polyketide synthase type I; PKS"  
/codon\_start=1  
/transl\_table=11  
/product="ShnA"  
/protein\_id="AAN63812.1"  
/db\_xref="GI:24637553"

/translation="MGVLADVHAGSEPOVALRSGAVLVRILARVADTDRASTGRRILD  
DVTALITGCTGALGALVARHLVVEHKIRSLVLSRRGPDAPGAADLDELTAIGARVR  
IVACIDAREAGELIASVPRDAPLTA VVHTAGVLDGVVGTALTPERIDAVLRPKADA  
ALVDELTRHLDVAFLVSSAAGTFGNPGQCNLAASNAYLDALARRRTAGLPATSV  
ANGWQDQGISDGLGVADQORWARWGLAAHSAQEGLELFDALRAYDAVLVAAXNFA  
GURAQASEFVHVLRLGLVRARRQAQSSREGGLAGLQAATPVQREQIILLDLVR  
EYAAVLYSTPRKVPDPAFQDVGTSLVAVELRNRLAGLIGLRIPASIAFDPTPRR  
MMRHLLAELCPEDGSEPADREDEIRRALATTLPSRFRELGLMEQILHLVAHPSGESAA  
APDTAPKQDAGPLIAEMDIDNLVKRAKEAKRP"

CDS

16813..22131  
/codon\_start=1  
/transl\_table=11  
/product="ShnB"  
/protein\_id="AAN63813.1"  
/db\_xref="GI:24637556"

/translation="MDARCONTRIAVASSKVEALRASIMENERLESEVQSTRDSLT  
EPIAIVGMACRFPFGVSSPEELWELLADRSVAEPPTNMGWDLNLDYDLPDCTT  
VYREGAFIHDAGFDAGFFGISQSETWMDPQORLMLETSWNEAFERAGIDPAMRCKN  
VYFAGMAAGQBYGTAFHSIPDELEGYVMTGSLVLSGRVSYTFYFGEPAVTIDTAC  
SSSLVALHMAAQSSGESSLALVGTGNVMTPTTAVLTARAGGLKGRCKAKFAASA  
DGTNWAEGVGLLIRLSDAVNRGVLVRATAVNODGASNLAAPNPSQQRVIR  
QALAGGLSPADVIVAEHGTCTALGDPTEAOALLTTYGNRAPGLPLWGSVKSLG  
HAGAAGVAGVIMWAMRHGVLPRTLHVDETPTEVDWSAGAVELLTEAHMPEVGRP  
RAGVSGFGASQTNHVLILEQATEPSTGMLPDEKARVIGDSVVPLVVSARGKAGLAGQ  
ARHLSGLTQRQDITDVLQSLVRSGRPLQNRVLAADREALAGLADAVARAESAP  
GVVTPFAESTVGRVTVLPFQQTQWAGMGAELIEASVFAARTECAVLDPLTGNLS  
LDVVRVQEGASLEDVVLQPVSWALMWLSALWEACGVVPAVVGHSIGEIAAACYA  
GALSLDARALMVHRSRIAEARIVGRGMSALTADVKVSIILEEWPGLIEIAAVNGPA  
SVVGTSELSELLEHLLARCADGIRARRIRGASINGASQSDILVDLHDFVALASVAGA  
SRVPLSYTVTGRHLDTEFDEVHEFRFNMATQVQDFPAIRSLVDGGRGVTEVSAHPVL  
TSSVQVDLADLEAGPAVVTGLTRDDGGPRFRFLASLAHLHTHGRVVRVDMWEALVGRGEGQ

Query Match 55.0%; Score 275.6; DB 1; Length 22131;  
Best Local Similarity 74.1%; Pred. No. 2.6e-20;  
Matches 366; Conservative 0; Mismatches 119; Indels 9; Gaps 1;  
PVDLPTAQRQRYWLETAESRGADEPIGLIEVANHPILGAVTEIPGSDGVLFTSLRLSL  
RTHFPLADHAGAGVNLPGAAFAVELAVRADEYGLVGLVLERPLVLPESGVQVR  
VWVPEDESHRTVQVHSRREAGSRGSTRVHSGRLVDEPDGQAEEDLTQWPPGATA  
VDPDAFAHYDIAEAGYHYGPAPQPMRAATWTEGEEVFAEVLSPESAGKADYGLHPA  
LIDAAHMTSLFRPLDSESPKALPFPVWRDLRHADGASTIRVHLPLAPDITLRLHLA  
DTSGTVPASVDSILVRPVPPELLRVSGGAQDMFRVAWEPI SVRSVDDDEKAVRVT  
AEDVRAAATAAPVLLDVAAGDRTDPDAARDLSGRVLEAQAALAPAFQDTVLIAL  
THSGAARDGDPVPLDVAATAAGLLRAAQSENVGRILLVDTGDEASARLDPVLAAG

4 GTCCCGGACGCGGAGGACGACGATCTCTGCTGAAGCTGGTCCGCGGCGACGCTTCGACG 63  
1903 GTCCCGGACGCGGAGGACGACGCTCTCTGCTCGGCTGGTCTGTGGCGACGCGCGGACG 1962  
64 GTGCTCGGCGGACGCGGCGGAGGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123  
1963 GTCTCTCGACACAGAGTGGGAGAGGATCGTCTCGCTCCAGTCTCTCCAGGAGGTGCGG 2022  
124 TTGCACTCGCTGGCGCGGCTCAACTCGCAACAGCTGCGACGCGGCGGCGGCGGCGGCGGCGG 183  
2023 TTGCACTCGCTGGCGCGGCTCAACTCGCAACAGCTGCGACGCGGCGGCGGCGGCGGCGGCGG 2082  
184 CTGCGCGGAGCGCTGATCTTTCGACTACCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 243  
2083 CTGCGCGGAGCGCTGATCTTTCGACTACCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2142  
244 GTCGAATCTCTCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303  
2143 TCCGAGCTGCTGACGGAACAGCGACGACCTGAGGCGGCGGAGGACGACCTGCGGCGGCGG 2202  
304 GTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363  
2203 GTGCTCGCACAGGTCTCCGCTCTCCGCGCTCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 2262  
364 GGCCTCGGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 423  
2263 AGCTTGGGCGGACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2322  
424 -----GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 474  
2323 GCGG 2382  
475 CAACGAGCGCTCGG 488  
2383 AAGCGGCTCTGG 2396

AF521897 6567 bp DNA linear BCT 10-MAR-2003  
Streptomyces hygroscopicus ansamycin biosynthesis gene cluster,  
partial sequence.  
AF521897  
AF521897.1 GI:28894473  
Streptomyces hygroscopicus  
Streptomyces hygroscopicus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 6567)  
Wang, Y. and Gao, Q.  
Cloning of ansamycin biosynthetic gene cluster, partial sequence  
from Streptomyces hygroscopicus 17997  
Unpublished  
2 (bases 1 to 6567)  
Wang, Y. and Gao, Q.  
Direct Submission  
Submitted (17-JUN-2002) Pathway Engineering Dept., Institute of  
Medicinal Biotechnology, 1# Beijing Tian Tan Xi Li, Beijing,  
Beijing 100050, China

FEATURES  
source  
gene  
CDS  
Location/Qualifiers  
1. .6567  
/organism="Streptomyces hygroscopicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1912"  
<1. .2414  
/gene="shnC"  
<1. .2414  
/gene="shnC"  
/codon\_start=3  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61220.1"  
/db\_xref="GI:28894474"  
/translation="EEDDNVRLRTVVGSTLRRDEGGPRRLTSMALFVRGVFVDMTK  
FLMAGHVDLPTVAFDRHVLWLODAATADDSDANDADADFWSAVEQTDADSLAGL  
LAPDSAGRLALRTVVPALADWGRSRRSSAERLRYAVTWRLDRVSRVPAGRWLA  
VLPPGPAETVTSRVAELIAELGAQGLDVVFFETVPSAFTRTGLTARLSDIRAYOP  
AGVLSLLALDGEQVDTIVARTLALVALQADGAVNGPLMCLTRGAVNTGIQDTAGEPG  
DAATWGLGAAALEHPDRWGLIDLPATADAHATQILVGLNGTAGDQLAVTTQAQPA  
RLVVRKASOTPADGWRPHGTIVLTGCAEALGHASLWARSAGRLIVTTQAQPA  
DAVTELOGLAAAGVETTVSCADARETLARLIAETPREQLTAVVHAADAPMTSAV  
ADTGHADLTVEFAGKVDTAWMLDELFTGDDAALDAFVVFSSIAIIMGGGCGGSGAA  
GADLDLVDLRRGRGLAATSTAWGALDGLGEMDEAAAQRLRRGVLPMAQVATF  
EQAAAREKAVTVADMDEAFIPAPTSARVSPADLPEAAAALSSQSPDAENGDTIS  
SLVDSLRDVPQAEQNRLLILVCGQAATVLGHSSGESIGLPQSFQEVGDSISLVNLR  
NSLWAGLRLPATLFDVPTPDVAVGFLRESELTETSDDLEGGEDDRLVLAQVPLS  
RLRAGLTLTLGLSDSVGSPVEXAPEAPAAAEADAEDRRDRGRPRKARSQQ  
PQQLT"  
2458. .5433  
/gene="shnd"  
2458. .5433  
/gene="shnd"  
/codon\_start=1  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61221.1"  
/db\_xref="GI:28894475"  
/translation="MTAPDEQIVDALRASLKENMRLOENQRLSSSSAPPIAIVSMAC  
RYAGTIRNPDLWRVNDQTDVTFSPENRGMDLEGIYHPDPDNFGTTYREGAFIHD  
ANLFDAGLFGISLEALAMEPQOLLEI CWEALERAGIDPHSVRGADIGVYAGLVHQ  
DYAPDLGEGTSLERALSAGGASRVAITLLEGPAVTVDTCGSTLVAVHVT  
QALRRGECAMALAGATVNSTPGFIFARQALAFDRCCKSYGAADGSSWAEAGV  
VILLERISDARNRHRVAVIRGSALNQDGASNLTPAGPAPORVIRKALENAGLITA  
DIDMVEGHTGTGLDPTTEAQTATYQDRPEGRPLMWGSKVSVIGHTQSSGAGL  
INAOALRGVMPATRVHDAPNPQVDSAGAVELITTEARWELGRPRRAGVSFGAS  
GTNAHMLEQAEPEAESPSAPALDGVPLVISAATASLTQARLGSFLEASGTV  
ALADVAALVTGRASLAQRAVVTDSPEALAGICALARGEDVRGVAGGVSQGDG  
KVLVFPQGGSPWGMGRLEECSEVFAARVGCBAVALERWYDMSIADVLRGDCPVEF  
FEREDVRQFASFAVMVGLAAVMSVGVVADAVVHSGGGEVAAACVXSLSLEDAVRV  
AVRSKTIISGLSGRGMASVGLSEEAVALQOQMDRVEIGAVNPSFSSVAITADTEAL  
DPAIETLBDQGVRRRIADYASHRSHVGAQVEILNEAFADIRSQAPVPLSTATGE  
WIREAGLDGSMYENLBSOVRFPAIDLLADGHTVFVESASHVPLVQPISEVVAGA  
EKRAVYLSGRHRHCGPRLTSMADLVFGRTHVDWGSVFLAAGADARRVDLPTAFDH  
KNYWMELAGTANDVASLGLSDHDLPGAVVPVETSGVLTCTSLRSLRTHPLWLAHVA  
GRCRACPHCFGCGGACG"  
5159. .>6567  
/gene="shng"  
5159. .>6567  
/gene="shng"  
/codon\_start=1  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AAO61222.1"  
/db\_xref="GI:28894476"  
/translation="MPAASTFRTPRLITTTGNSWVPVPTMSPRSAACRGLIIRCWVRW  
FRERRAECCVRGRCFHRTIRITLMAGVVLVPGTALVELVVRAGDKVCGCTLEL  
VIETPLVPAQSMRVQFAVGGPEENGSAVSAYSDDDGRTGIDGTRWRRAGEVTF  
AAVADGDFDFTVPVGAERVSFADYFTEEMAGRGYGYFAGFGLRGWRHGEVVF  
AEVALPDEOHGEASRGLHALLDALQSLGVRPADGVMDRVPFAMWGLRLHAAGS  
ELURVTVPSPDAVSLOADAGGPPVLSLESLVRAVDVDEQLDRMATDGRDALFEVD  
WSELPAPASSVSLPSPASVASEDVTDLADAAVPAVAVLEAVGEGGSHDALALTVR  
VLEWVQAAAGLAESRLVVVTRGAVPVGEGGNVADPAGAVGLVRAAQAENPDR1



```
/code="orf2"
/transl_table=1
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEOLVRLAKVSAKENARLRRENASLOEGAEVPAIVGMACRL
FGGVSPESLWELMAGDQVSPFPDTRCWDLDGLYDPPDHLGTAYVREGGFLDRAG
AFDPPFGISPREALAMPDQORLLLETTWEVLERAGIDPTTLRGRTGVTGFTIGLQDY
VTRLDDPFAELGSLGNGSAASVAGRSVYTFEGEPVAVSLDTACSSSLVALHLVRP
ALRLLARVAGVTWVSVHPYFSPQRGLAFDARCKFSQADGTVLSEGVGLLV
ERLSDALRGVPLVAVRTAVNQDASNGLTAPTQSQORVIRDALTDARLTADVDD
LLEAGHTGRIGDTEAQAALATYQGRDQPVWLSGLKSNIGHTQAAAGVAGVITKVM
AMRGVMPRTLVSEPTNVDMSAGEVRLTWARDWPSDQRPRAAVSSFGISGTA
HVVLQCAAPAPDVPVTPAGVPVWLSAHTGPAISALAEHLAHEDEAPLGDA
HYLNTATLPHRAVVGSTREEFVAALRAVPTPRRGLPAVMTFGQSQRPMGRQ
LLEHFFVFAQLDEVPVATGLREVMEDPDQAEITLQRTDHAQIALFAPETALRYME
SGLGRPDMVCGHSVGEITAAHNGHATVLSGRDTHLHIAQTMLNKLNVSHAFSP
HTLQPHLDNHDTISIAAINGPHATVLSGRDTHLHIAQTMLNKLNVSHAFSP
MOPILQPFTHLTHLTHPPTLISMLTATPHDPTHTWTOHIATPYRYTDLHLH
HNGITYLSEIGDPTTLTALARTLPTTLTHLPTTRNNEVSTNEALGRVFSVSHV
DMRALTGRRSLTPFPQRDPMHDAAGGAEEVAGLGTDRHLLGAVTVVADT
GELLISGRUSTHPLWDTHTVNGTVIPEGTALIDLALHAAETHHTVDELIVHTPL
ALHTHPTSGTGAEDTGHRTPLALHSRTTGTWHTTGLTSLNTHPAAGLPTWSPT
DAHQIDLTAAYQAADTGLHGPFAFQGLHTLYORDNTLFAELPEAAGLPTGHIHP
ALLDALHAPAAARDESGIQLPESWAGVTVHSGAHLVRLDTHDGTALRLAT
DTGQPVITSLSTRPLHDPDQHLAAQPEINHLITLGNSTPFLITDTTGTGFERIVPT
AGDILLADTHRTIAQTLRIQOHLAADTTLPLVVEATHDDLAGRVMLRTAQTETHP
GRILLDNDHPTSDALASVAVSGRQARIRDAITLRLTSLSDVTRASTERRPDL
SMGVLVTGCTGGLSLIARHIVTEHGVLDLVLVSQGDADGATBLTNLOQHGARI
RTISCDLTORTALTLLIDBIGLTVGVVHTAGALADTTIDHLPDNTLTLTKANPAW
WHEHTODHDLALFVFFSVAGVILGNAGQANAAANSELDALITRRRQGLPETSIAW
GHWQREGMTHLQADDRMTRAGHGTDAEGTTLFTDLDRGGAAPVAKHLPT
LNRADTVAVLGLKRVAPPAARQVTPAESWADRIAALPEERDRAVLDVROQSR
VLGHSDADRIDPGAFKDFDPSLTAVEFRNRTVATGLRLPATLIFDPTFNALADH
LIAQLQGLGGSGSPVVRDLERFERQLFDRPDDEVTAGTAIVRLRSILAHLDAPAT
APAARTASQVLDLLEDAQSSELLSFIDKEFRASN"
14964..21266
/notes="orf3"
/codon_start=1
/transl_table=1
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="MADSKTLDLYLKNVTADLYRTRERLAEVESASAEPIAIGMACR
FPGVTSPEALMELVSSGDTLADSGPDDDRGDMWBGFFDSIPAQEGTSYMRGEGFLTA
TGFDPAFFGISPREAVAMPDQORLLLETTWEVLERAGIDPTTLRGRTGVTGFTGLMANS
YTRVVRIRIPELESVAGNEAASVASGRSVYTFEGEPAMSVDTACSSSLTALHLAVR
ALRSGEPLAVAGGVTPATPRIFSDLSRQGLARDGRCKAFSTADGTGLSEGVGLL
LYERLSDAORLGHSLGVVGRGTAVNQDASNGLTAPNGPAQORVILKAALADARLTAD
IDLLIENHGTGKLDPIEAQALATYQGRDQPVWLSGLKSNIGHSQAAAGVAGIYKI
VWAMRHGVMPKTLIHIDEPTPEVMSAGTVELLAEARDWPFVDERPRRPGVSSFGSGT
NAHILEQAPPAPEPTSEAEVPESEAGSPGSGVLELGPSPVWLSAQNETALKD
LAEQATVDGSDVRDIHAHSLISTRAALPHRAVVVSGSTREEFATALEQITGHTPDGKTV
LVFPGGSGWGMGQRLWTSPLYRDTITDIALSEYIDWSVDTLTSQPGAPWDR
VDVIOPLFTWMTALATLQHQHGLTPDALIGHSQGEIAAAYTAGALTLODAARITLIR
SKALLPLTGHGTWISVLAPQIOTIADPDHWNIRIWIHAHAPNTITITIGTQALH
LQQLSHRIMRWQLPGVDFAGHSHIDQLHQLKILAPITPQKPTIFPFTDNTW
IHTTLADYVYRNLRFQFOPAINHLNNGYTYIETSPHPTLPSTQETNPNTT
ITDLSNQNDTHALITAGHANTHGPITWHTLIPPTKTPPLTPYTFORTHEWLNKTI
ADAVESAGLSRTHDPLGLATLHSGETVLNGLSLTHHGWIAOHVNGTPLLPGT
ALVDMALHAGDHTDTLLELIHTPTLTHTPTIQLTTPPTDHTDNTLITITRAP
EQPWIOHAGTLTRANTASSPETIWPPTNAEPIIDGTYVQRIATLGLSEYGAQGLH
ALWRDSDNRLLADVLEPDTINDTDCDGDYGIHPALLDHALHALVAVSDGEIRL
PFAWTGTLHATKARLVELDIDDDGTLSLSATDTSQGVVTVSDVSTRPIDPAQLR
APRQDIDLALDMLDLAGAAVWGLRTAQTHPGRILIDTNDHPTSDALASVVAS
AADTPTPLVVEATHDWSAAPPFEQVEVPKVGDDVLADTHRTIAQTLTRVQOHL
GEQPARIRGALTLPRITAVTDALTTPADADAMHLDTSERGTLNENLITGSDATIR
NLLPGEVRVAVRAAGLNFRDVLIALGMYFGEATMGSEAGIVLIDIGDITLTPGDRV
TGLFFPNALANTAVADRMMTIPIPASWNSNTDAAFFIVILYATYGLHDLAALKEPGERVL
IHAGTGCGVGTAAIQLAQHGAEVAFATHTKPHLRAMGLDHDHIIASRTPIPFTH
TTTQGGQGHVILNLDLTDASLRLLTPNGRFIEMGKTIDVRTEVARTHPQVTVRA
FDIADGAPDRIGAMVELTGLPERGALRGLPVRCDVRESVDAPFRMAQARHIGKV
```

```
CDS
21363..27176
/notes="orf4"
/codon_start=1
/transl_table=1
/product="PKS module 4"
/protein_id="AAC38064.1"
/db_xref="GI:2624951"
/translation="MANEDDFLHYLKKAAADLRDARQIQOLEAREDRPIAIVSMACR
LFGGVSPESLWELMAGDQVSPFPDTRCWDLDGLYDPPDHLGTAYVREGGFLDRAG
AFDPPFGISPREALAMPDQORLLLETTWEVLERAGIDPTTLRGRTGVTGFTIGLQDY
VTRLDDPFAELGSLGNGSAASVAGRSVYTFEGEPVAVSLDTACSSSLVALHLVRP
ALRLLARVAGVTWVSVHPYFSPQRGLAFDARCKFSQADGTVLSEGVGLLV
ERLSDALRGVPLVAVRTAVNQDASNGLTAPTQSQORVIRDALTDARLTADVDD
LLEAGHTGRIGDTEAQAALATYQGRDQPVWLSGLKSNIGHTQAAAGVAGVITKVM
AMRGVMPRTLVSEPTNVDMSAGEVRLTWARDWPSDQRPRAAVSSFGISGTA
HVVLQCAAPAPDVPVTPAGVPVWLSAHTGPAISALAEHLAHEDEAPLGDA
HYLNTATLPHRAVVGSTREEFVAALRAVPTPRRGLPAVMTFGQSQRPMGRQ
LLEHFFVFAQLDEVPVATGLREVMEDPDQAEITLQRTDHAQIALFAPETALRYME
SGLGRPDMVCGHSVGEITAAHNGHATVLSGRDTHLHIAQTMLNKLNVSHAFSP
HTLQPHLDNHDTISIAAINGPHATVLSGRDTHLHIAQTMLNKLNVSHAFSP
MOPILQPFTHLTHLTHPPTLISMLTATPHDPTHTWTOHIATPYRYTDLHLH
HNGITYLSEIGDPTTLTALARTLPTTLTHLPTTRNNEVSTNEALGRVFSVSHV
DMRALTGRRSLTPFPQRDPMHDAAGGAEEVAGLGTDRHLLGAVTVVADT
GELLISGRUSTHPLWDTHTVNGTVIPEGTALIDLALHAAETHHTVDELIVHTPL
ALHTHPTSGTGAEDTGHRTPLALHSRTTGTWHTTGLTSLNTHPAAGLPTWSPT
DAHQIDLTAAYQAADTGLHGPFAFQGLHTLYORDNTLFAELPEAAGLPTGHIHP
ALLDALHAPAAARDESGIQLPESWAGVTVHSGAHLVRLDTHDGTALRLAT
DTGQPVITSLSTRPLHDPDQHLAAQPEINHLITLGNSTPFLITDTTGTGFERIVPT
AGDILLADTHRTIAQTLRIQOHLAADTTLPLVVEATHDDLAGRVMLRTAQTETHP
GRILLDNDHPTSDALASVAVSGRQARIRDAITLRLTSLSDVTRASTERRPDL
SMGVLVTGCTGGLSLIARHIVTEHGVLDLVLVSQGDADGATBLTNLOQHGARI
RTISCDLTORTALTLLIDBIGLTVGVVHTAGALADTTIDHLPDNTLTLTKANPAW
WHEHTODHDLALFVFFSVAGVILGNAGQANAAANSELDALITRRRQGLPETSIAW
GHWQREGMTHLQADDRMTRAGHGTDAEGTTLFTDLDRGGAAPVAKHLPT
LNRADTVAVLGLKRVAPPAARQVTPAESWADRIAALPEERDRAVLDVROQSR
VLGHSDADRIDPGAFKDFDPSLTAVEFRNRTVATGLRLPATLIFDPTFNALADH
LIAQLQGLGGSGSPVVRDLERFERQLFDRPDDEVTAGTAIVRLRSILAHLDAPAT
APAARTASQVLDLLEDAQSSELLSFIDKEFRASN"
14964..21266
/notes="orf3"
/codon_start=1
/transl_table=1
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="MADSKTLDLYLKNVTADLYRTRERLAEVESASAEPIAIGMACR
FPGVTSPEALMELVSSGDTLADSGPDDDRGDMWBGFFDSIPAQEGTSYMRGEGFLTA
TGFDPAFFGISPREAVAMPDQORLLLETTWEVLERAGIDPTTLRGRTGVTGFTGLMANS
YTRVVRIRIPELESVAGNEAASVASGRSVYTFEGEPAMSVDTACSSSLTALHLAVR
ALRSGEPLAVAGGVTPATPRIFSDLSRQGLARDGRCKAFSTADGTGLSEGVGLL
LYERLSDAORLGHSLGVVGRGTAVNQDASNGLTAPNGPAQORVILKAALADARLTAD
IDLLIENHGTGKLDPIEAQALATYQGRDQPVWLSGLKSNIGHSQAAAGVAGIYKI
VWAMRHGVMPKTLIHIDEPTPEVMSAGTVELLAEARDWPFVDERPRRPGVSSFGSGT
NAHILEQAPPAPEPTSEAEVPESEAGSPGSGVLELGPSPVWLSAQNETALKD
LAEQATVDGSDVRDIHAHSLISTRAALPHRAVVVSGSTREEFATALEQITGHTPDGKTV
LVFPGGSGWGMGQRLWTSPLYRDTITDIALSEYIDWSVDTLTSQPGAPWDR
VDVIOPLFTWMTALATLQHQHGLTPDALIGHSQGEIAAAYTAGALTLODAARITLIR
SKALLPLTGHGTWISVLAPQIOTIADPDHWNIRIWIHAHAPNTITITIGTQALH
LQQLSHRIMRWQLPGVDFAGHSHIDQLHQLKILAPITPQKPTIFPFTDNTW
IHTTLADYVYRNLRFQFOPAINHLNNGYTYIETSPHPTLPSTQETNPNTT
ITDLSNQNDTHALITAGHANTHGPITWHTLIPPTKTPPLTPYTFORTHEWLNKTI
ADAVESAGLSRTHDPLGLATLHSGETVLNGLSLTHHGWIAOHVNGTPLLPGT
ALVDMALHAGDHTDTLLELIHTPTLTHTPTIQLTTPPTDHTDNTLITITRAP
EQPWIOHAGTLTRANTASSPETIWPPTNAEPIIDGTYVQRIATLGLSEYGAQGLH
ALWRDSDNRLLADVLEPDTINDTDCDGDYGIHPALLDHALHALVAVSDGEIRL
PFAWTGTLHATKARLVELDIDDDGTLSLSATDTSQGVVTVSDVSTRPIDPAQLR
APRQDIDLALDMLDLAGAAVWGLRTAQTHPGRILIDTNDHPTSDALASVVAS
AADTPTPLVVEATHDWSAAPPFEQVEVPKVGDDVLADTHRTIAQTLTRVQOHL
GEQPARIRGALTLPRITAVTDALTTPADADAMHLDTSERGTLNENLITGSDATIR
NLLPGEVRVAVRAAGLNFRDVLIALGMYFGEATMGSEAGIVLIDIGDITLTPGDRV
TGLFFPNALANTAVADRMMTIPIPASWNSNTDAAFFIVILYATYGLHDLAALKEPGERVL
IHAGTGCGVGTAAIQLAQHGAEVAFATHTKPHLRAMGLDHDHIIASRTPIPFTH
TTTQGGQGHVILNLDLTDASLRLLTPNGRFIEMGKTIDVRTEVARTHPQVTVRA
FDIADGAPDRIGAMVELTGLPERGALRGLPVRCDVRESVDAPFRMAQARHIGKV
```

```
Query Match 32.8%; Score 164.4; DB 1; Length 32870;
Best Local Similarity 60.1%; Pred. No. 5.8e-09;
Matches 273; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 12 CGCGGAGCAGACCGCATCCTGCTGAAGCTGTGTCGCGGCGCAGCGTTCGACGGTCTCGG 71
Db 14415 CGAGGAGCGGCGACCGGCGGTGCTGACCTGTGTCGCGGCGCAGCGTTCGCGGCGG 14474
QY 72 CCACAGCGCGCGGAGGCGATCGGCGCGCGCCAGCGGCGTTCAGGAGGTTCGAGTTC 131
Db 14475 GCATAGCAGCGCGGACCGGATCGACCCGCGCGGCGGTTCAGGACACCGGTTTCGACTC 14534
QY 132 GTTGCGCGCGGTCAACCTCCGCAACAGCCTGACCGCGCGCCACCGGCGTTCGCGTCCCGC 191
Db 14535 GCTCACCGCGTTCGAGTTCGCGACCGCGTTCACCGCGCGGCGTTCGCGTCCCGC 14594
QY 192 GAGCTGATCTTCGACTACCCACCGCGGAGGCGTGTGTCGCTACCTGCGGCTCGAAT 251
Db 14595 CACCTGATCTTCGACTACCCACCGCGGAGGCGTGTGTCGCTACCTGCGGCTCGAAT 14654
QY 252 CCTCGGAGGCGCGACGCGGCTGCGCGGCGGGAAGCAGCCTCCGCGAGTTCCTCGC 311
Db 14655 CGGCGCATTCGCGGCGCGGCGGCTGCTGCTCCCGGAGGTTCGCGGACCTGAGCGTTT 14714
QY 312 GGCGTTCGCGTTCGCGCGGTTCGAGGAGGCGGCGGTTCGAGCAGCGTTCGCGCTCGC 371
Db 14715 GGAACGCGATTCGTCGACACCGCGCGGCGGAGGAGTGGGCGCGGATCGCGTCCG 14774
QY 372 CGACACCGGCGCGAAGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 431
Db 14775 GTTGCGCTCCATCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14834
QY 432 CGACGCGAGACTGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 465
Db 14835 CAGCGCGGCGGCTTCGACGACGCTGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14868
```

```
RESULT 7
BD217187
LOCUS BD217187 4689 bp DNA linear PAT 17-JUL-2003
DEFINITION DNA encoding methymycin and pikromycin.
ACCESSION BD217187
VERSION BD217187.1 GI:33026957
KEYWORDS JP 2002536959-A/17,
Streptomyces venezuelae
```

```

gene cluster, complete sequence.
AF079138
AF079138.1 GI:3808326
.
Streptomyces venezuelae
Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 37948)
Xue.Y., Zhao.L., Liu.H.W. and Sherman.D.H.
A gene cluster for macrolide antibiotic biosynthesis in
Streptomyces venezuelae: architecture of metabolic diversity
Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
98445333
9770448
2 (bases 1 to 37948)
Xue.Y., Wilson.D., Zhuo.L. and Sherman.D.H.
Direct Submission
Submitted (17-JUL-1998) Department of Microbiology, University of
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
3 (bases 1 to 37948)
Xue.Y., Wilson.D., Zhuo.L. and Sherman.D.H.
Direct Submission
Submitted (29-OCT-1998) Department of Microbiology, University of
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
REMARK
COMMENT
FEATURES
source
1. .37948
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
154. .1122
/genes="pikR2"
154. .1122
/genes="pikR2"
/notes="mIs type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR2"
/protein_id="AAC69327.1"
/db_xref="GI:3800832"
/translational="MAFSPQGRHELGONFLVDRSVIDEIDGIVARTKGPILGIPGD
GALTPLSHGRPTIAVELDGRRAQRLGCHTVVHHDELQYPLPRNPHVVGVN
PFLHTAIWRLLDADHWHTAVLLVQWEVARRAGVGGSTLLTAGWAPWTFDLHVR
PARAFPMQPDGVGLAIRRSAPLVQKVQYDFRVQVPTKGNGKELIRRTKGRIS
QRLATLWRNELSPHALPKFGQWASIMELTGTADGFDGTAGGGAAGSHGAAR
VGAGHPGGRVSRGVPQARSGRCHAVRSSTGTEPRWGRGSAES"
complement(1283. .2293)
/genes="pikR1"
complement(1283. .2293)
/genes="pikR1"
/notes="mIs type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR1"
/protein_id="AAC69328.1"
/db_xref="GI:3800833"
/translational="MAMRDSIPRRADRTLRRRLGONFLDQDRAVRNLVTHVGDGGRN
VLEIGPKGKATIELVRSFDTVTVMDDPWAHVRRKPEGRVTYFQGFDFDRI PR
DITTVGVNPFGITTOILASLSTWQSAALVQWEVARRAGVGGSTLLTAGWAPW
YEFADHVRVASFRPMRVDDGVLTIRRPQPLPESARAFQNFABVTFQPGRL
AETLRHPIPRYRLADRHGIDGGLPKDLTLTQWIALFQASQPSYAPGAPGTRMP
QCKGAGGRDYDSETSAAPVGRSRYGPTRGCEPCAPRAQVQRTKGRQCARSSYGRRT
GR"
2912. .16753
/genes="pikAI"
2912. .16753
/genes="pikAI"
gene
CDS

```

```

/notes="multifunctional polyketide synthase; contains a
loading module, Pik module 1 and module 2"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase PikAI"
/protein_id="AAC69329.1"
/db_xref="GI:3800834"
/transl_table="MSRAGITRTTCARTPTVTCRGAAMWTCHEVRVRRLPGAGPDHAEH
SFSRATGCDVAAELIRGEMSTVSKSEEEFVSVDNAGSAGHSTAEPAVAVGICSRVPG
ARDPREFWELAAAGQAVATVPDARNWAGDIPDRSAPAGSRTNKGWGFIEDVDRFDA
AFFGCSIPREAEVDPOORLALGLWEALERAGIDPSLITGRTRGVFAAGIWDYATLTK
HQCGAAITHTVITGLRGIITANRLSYTLGRGPMVVDSCQSSSIVAHVLAHCSLRR
GESLALAGVSLVLPDSIIGAGFCGLSGDRAYTFDARANGYVRGEGGQAPVYVLR
LSRAVADGDPVLAVIRSGVAGGAAOGMTTPDAQAEAVLREAHERRAGTAPADVYV
ELHGTGTPVGDIIEAALGAALGTCRPAQQLLVGSVKTNIHLEBAGIAGLIKAVL
AVRGRAIPASLYNETPNAIPFEELNLRKNTYELPWEPEHGDQRMVGVSVFGMGNTD
AHVLEEAPGCGRCSTGVSTVGGSVGGVPMVVSAAKSAALDAQIETLRAASGRN
RTDGVADAGVADAGAVAGVARVLAGRAQOEHRVAVVGGPDDLAALAAPEGLVRG
VASGVGVAFVFCGQTCWAGCAELLDSSAVFAAAMAECEAALSPYVDMWLEAVVRO
APGAPTLERUDVQVPTFANVSLARVWQHGVTPQAVVGHSGOGEIAAAVYAGALSLD
DAARVTVRSKSIASHLAGKGMLSLASEDAVLERLAGPGLGSVAANGTATVVSIG
DPVQIETELARACEADGVRAVPIVDYASHSRQVEIIESELAELVLAGSPQAPRPFPS
TLEGATITPEYLDGQVYRNIRHRVGFAPAVETLATDEGTFHFEVSEHPVLTWALPG
TVTGLATLRNDGQDLVSLAEAWANGLADWSPLLPSSATGHSLDLPYAFOTERH
WLGETALAPAGEACTSLGQCTVLRTEARPAELDRDEOLRVIIDKVRACQTAQVLYGATGG
QIEVDTRPREAGCTSLGQCTVLRTEARPAELDRDEOLRVIIDKVRACQTAQVLYGATGG
AANAPGAPAPVAAAGAVDEPVALVGMACHRLPGVSPEDLMRLVAGGGBALSEFPQ
DRMGDEGLHDPDDEHPTGSVYRQGGFIENVAGDGAFFGSIAPREALANDPQORLILLE
TSMWEADAGIDPTSLRQGRQVFTGAMTHEYFGPSLRDGLDGLYLTGNTAYMSG
RVSYTLGLEGPALTVDYTCSSSILVALHVALQALRKEVDMALAGVAMPTPGMFVEF
SRQGLAGDGRSKAFAGADGTSWEGVLLVERLSDARNAGHOVLAVVRGSAALNQD
GASNGLTAPNGSPQOVRIRALADARLTSDVDVVEAGHGTGRLGDPTEAOLATYTG
QGRDDEQLRLGSKNIGHTQAAAGVSGVIGKGNANHVLEAPVVEGASVVERSVG
AGAVELLTEADWPEKQDGLRRAAVSFGISGNTANHVLEAPVVEGASVVERSVG
QSAFVGTPVWVSAAKSAALDAQIETLRAASGRNRTDADAGVADAGVAHVLDAGR
AQFHRVALVAGADLDVQALADPDGLIRGTASGVGVAFVPGQGTQWAGCAELLD
SSAVFAAAMAECEAALSPYVDMWLEAVVROAPGAPTLERUDVQVPTFANVSLARV
QHGVTPQAVVGHSGOGEIAAAVYAGALVLDAAVAVVTVLSKSIASHLAGKGMLSLAL
NEDAVRTPQVGHSGOGEIAAAVYAGALVLDAAVAVVTVLSKSIASHLAGKGMLSLAL
HSRQVEIIESELAQVLAGLSPQAPRPFPSLTLEGTWIITEPVDIGTYVYRNLRHVGFA
PAIETVAYIDEGTFHFEVSEHPVLTWALPTEVTLGLTRREGQGERLVTSIAEAWN
GLPVANTILLPATASRPGPLTVFAQEARWYLENTPAALATGDMWYRIDWRLPAAG
SERLTCGLRWALTDEDSQAQAAVLTALVDAGAKVEVLTAGADDDRALAARLTALT
TGDGTGTVSYLLDGLVQVAVQALGDAGIKAPLWSVTCGASVGRLOITPADPRAML
WGLGRVLEHPEHFWAGLVLDLPQDPAALAHHLVTLASGATGEDQIARTIRGLHAPRL
ARAPLHRRPTQVDPHGTVLITGGTALGSHAAWMAHGAHLLVSRSGEQAPGA
TQTLAELTAGARVITIAACDVADPHAMRTLDLDAIETPLTAVVHTAGALDDGIYDTL
LAARRRATGRSAVAVWGPWPGMGMAAGDVAEIRNKGVPGMDELAALAESALGR
DETAITVADLDWDRFLAYVSGRPOPLVEPVERIIDARSAVSGOGSSAQGANP
LAERLAAAPGERTELLGLVRAQAAVILRMKSPEDVAADRAFQIDGDSLAGVELRN
AFSADADGFGAEGVGLLILVERSDARNNGHPVLVVRGTAVNQDASNGLTAPNGPS
QORVTRQALADRALPGDIDAVETHTGTSIDGPIEAQGLQATYIKRERAPRLAIGS
VKSNTITQQAAGAGIIGKNVLMHRHGLTKTHADEPSPHVDWANSGLALVTEPIDW
PAGTGPRRAAVSFGISGNTANHVLEAPVVEGASVVERSVGDEPVEVETVAMAGTAGE
VAGSEASAPAPAGSREASLPGLHPWLISAKDEQSLRQAAALHAWLESAPADLSDA
DGPALRDVGGYTLATSRTPAFRAAVTAADRDGFLDGLATLAQGGTSAHVHLDITARDG
TTAFLFTGGVGRPGHRELIDRHPVFRALDDEIICHLDGLHLEPLLDVWFAEGSAAE
AALLDETRYTQCALFALEVALFVLSWGRWEPAAALIGHVSGEIAAHVAVGVSFLADAG
RLVFAARQLWQELPAGGAMLAVOAEDIEIRVMLETEERYAGRLDVAANGVPEAAVLS
DADAAREAYVSGLGRRTALRVSHAFSHAMDGMDFGFRACVLETFEVRPRLSTVVS
NVTYGLAGDPDLDCEPVYVHVHVGTVFRLDGVRLDGLVTRCLRIUGDPGLVTAMAAD
GLADTPADSAAGSPVSGPAGSDSAGALRPRPLVALLRRKRSSETVYADALGPAH
AHGTGPDHMAWFAAGAHVRLDPTYSFRDRYVLDPAADAVTDTHAGLGLGTADHPLV
VGSILPDRDLGLLTLGRLSRTHPLADHVLGSLVLLPGAAVVELAAHAAESAGRDV
RELTLLEPLPHGGVGLRVTVGAPGPGESAGDGPVPSLHSGRLADAPAGTAW
CHATGILTDRPELVPDRAAMWPPQGAEEVPLDGLYERLDNGNLAGFPLFOGLNAV
WRYEVEVPADIALPATNATATANGGGSAAAAPYGIHPALLDASLHAIVGELVDE

```

```

Query Match      31.7%; Score 158.6; DB 1; Length 37948;
Best Local Similarity 59.0%; Pred. No. 2.2e-08;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY      1  GCGGTCCTCCGCGCGGCGAGAACCGCATCTCTGCTGAACTCGTCTCCGCGCGCCAGCTTCG 60
Db      32343 GCGCTCCCGCGCGCGCGCGCGGCGCTCTCTCACCTCGTCTCGTACCCACCGCGCG 32402

QY      61  AGCGTGCTCGGCGCACAGCGCGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 120
Db      32403 GCGGTACTCGGCGATCTCTCTCCCGGCGCGGCGCGGCGCGGCGGCGGCGGCGGCGG 32462

QY      121 GCGTTCGACTCGCTGCGCGCGGTGAACCTTCGCAACAGAGCTTCGACCGCGCGCCCGGCGCTG 180
Db      32463 GCGTTCGACTCGCTGAGCGCGCGGTGACGCTCCGCAACAGAGCTTCCTCACCGGTGTCGCGAAC 32522

QY      181 CCGGTCCCGCGCGAGCGGTGATCTTCGACTACCCACCGCGCGCGGCGGTGCTCGGCTACCTG 240
Db      32523 AGGCTCCCGCGCGAGCGGTCTTCGACCAACCGCGCGCGGCGGCGGCGGCGGCGGCGG 32582

```









```
Db 29735 GACCTGTTGGAGCGGAGCGTGCATGACGACCTGGAGCGGAGCCCTGATCGG 29794
Qy 478 CGAGCCCTCGGCG 490
Db 29795 ATGGCTCTCGGCC 29807

RESULT 13
AX697993
LOCUS AX697993 4725 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 17 from Patent WO03010193.
ACCESSION AX697993
VERSION AX697993.1 GI:29499042
KEYWORDS
SOURCE Micromonospora carbonacea
ORGANISM Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1
AUTHORS Farnet, C.M., Staffa, A. and Yang, X.
TITLE Genes and proteins for the biosynthesis of rosaramicin
JOURNAL Patent: WO 03010193-A 17 06-FEB-2003;
Ecopia Biosciences Inc. (CA)
FEATURES
source 1. .4725
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
ORIGIN
Query Match 31.1%; Score 155.8; DB 6; Length 4725;
Best Local Similarity 62.2%; Pred. No. 8.7e-08;
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
Qy 12 GCGGAGCAGAACCGCATCTCTGTAAGCTGGTCCGGGCGCACGCTTCGAGGTCGCTCGG 71
Db 4254 CGACGAAACGCGCGGGGCTGCTCGACCTGTAGCGCGGCGAGGTGCGCGCTCTCGG 4313
Qy 72 CCACAGCGGCGCGAGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
Db 4314 CCACCGGGGCGCGAGCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4373
Qy 132 GCTGCGCGCGGTCAACCTCGCGCAACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191
Db 4374 GCTGACCGCGCTGCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4433
Qy 192 GACGCTGATCTTCGACTACCCACCGGAGCGCTGCTGCGGCTACCTGCGCGCTCGAACT 251
Db 4434 CACCTCTGCTTCGACCAACCCACCGGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCT 4493
Qy 252 CTTGGGAGCGCGAGCGGCTGAGCGGCGGCGGAGACGACCTCGCGGAGTCTCTCGC 311
Db 4494 CGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4550
Qy 312 GCGCGTGCCTGTTGCGCCCGGCTCAAGAGCGCGGCGCTGCTGACACGCTGCTCGGCGCTCGC 371
Db 4551 CGACTGCGCTGCGCGCGCGCTGCGGAGCGCGGCTACTGAGCGGCTGCTTGGCTTGC 4610
Qy 372 CGACACCGGCGACCGAACCGGCGACCGAGCGCGGACCGAGACCGAGCGGCGCGCGCG 431
Db 4611 GGGGCTGAGCGCGGAGCGGCTGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 4670
Qy 432 CGA 434
Db 4671 CGA 4673

RESULT 14
AX697977
LOCUS AX697977 60196 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 1 from Patent WO03010193.
ACCESSION AX697977
```

```
VERSION AX697977.1 GI:29499034
KEYWORDS Micromonospora carbonacea
SOURCE Micromonospora carbonacea
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1
AUTHORS Farnet, C.M., Staffa, A. and Yang, X.
TITLE Genes and proteins for the biosynthesis of rosaramicin
JOURNAL Patent: WO 03010193-A 1 06-FEB-2003;
Ecopia Biosciences Inc. (CA)
FEATURES
source 1. .60196
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
ORIGIN
Query Match 31.1%; Score 155.8; DB 6; Length 60196;
Best Local Similarity 62.2%; Pred. No. 3.6e-08;
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
Qy 12 GCGGAGCAGAACCGCATCTCTGTAAGCTGGTCCGGGCGCACGCTTCGAGGTCGCTCGG 71
Db 40545 CGACGAAACGCGCGGGGCTGCTCGACCTGTACGCGCGCGAGGTGCGCGCGCTCTCGG 40604
Qy 72 CCACAGCGGCGCGAAGGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
Db 40605 CCACCGGGGCGCGAGCAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40664
Qy 132 GCTGCGCGCGGTCAACCTCGCGCAACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191
Db 40665 GCTGACCGCGCTGCGACCTGGCCAAAGCGCTCAGGCGCGCGCTGCGCGCTGTCGCG 40724
Qy 192 GACGCTGATCTTCGACTACCCACCGGAGCGCTGCTGCGGCTACCTGCGCGCTCGAACT 251
Db 40725 CACCTCTGCTTTCGACCAACCCACCGGAGCGGCTGCGCGAGCACCTTGGCGCGGCTGCT 40784
Qy 252 CTTGGGAGCGCGACGACGCGCTGAGCGGCGGAGAGACGACCTCGCGGAGTCTCTCGC 311
Db 40785 CGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40841
Qy 312 GCGCGTGCCTGTTCCCGCGGTTCAAGAGGCGGCGGCTGCTGACACGCTGCTCGGCGCTCGC 371
Db 40842 CGACTGCGCTGCGCGCGCTGCGGAGCGCGCGCTACTGAGCGGCTGCTTGGCTTGC 40901
Qy 372 CGACACCGGCGACCGAACCGGCGACGCGCGCGAGACGACCGAGCGGCGCGCGCGCGCG 431
Db 40902 GGGGCTGAGCGCGCGCTGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40961
Qy 432 CGA 434
Db 40962 CGA 40964

RESULT 15
AX607061/c
LOCUS AX607061 15970 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 147 of the
complete genome.
ACCESSION AE007061 AE000516
VERSION AE007061.1 GI:13881778
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 15970)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, T., Khouri, H.,
```

Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 15970)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.P., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Ufferback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-Apr-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES  
source  
1..15970  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
70..726  
/gene="WT2106"  
/gene="WT2106"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK46385.1"  
/db\_xref="GI:13881779"  
PAAAGPITIVSWIAGHPVDPAAHVATRDVTYQLGDDVAFSSGTVACMTDARH  
TSGLTAVADNPPREPTATGEWGWDFDGLHQVSARADPGFVINGFELA  
NGDTLSIGDYRCRSYQAGLFCVNYAHQSAVFASAGIBPFCKLPAPPDPGVGVGFC  
"  
complement (763..3327)  
/gene="WT2107"  
complement (763..3327)  
/gene="WT2107"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK46386.1"  
/db\_xref="GI:13881780"  
/translation="WRIAVTGASVGLRGLTARLLSQGHEVVGIARHRPDSWPSADP  
IAADRDATAVESMTGADVVAHCAVGRNDHINI DGTANVLKMAETGTGRIYFTS  
SQHQPRVEQMLADCGLEWAVRCAITFGNVDNMQRLPALPVLPGYADRVVQVHS  
DDAORLLVRLALDVTIDSGPVNLAAPGELTFRIIAALGRPMVPTGSPVLRVTGFAE  
LELLHSAPLMDVTLARDWGFQPAWNAECLDFTLAVRGRIGLKRTFSLPRLANI  
QDLPAVDSPADGAVAPRLAGPGANGEEFTPIDRPPTYLATNLSEALPGPSPSAS  
VTVGRAGGVGIARLRPSGVIQREIAMRTVAFAHRLYGAITSAHFWAATVPFAKP  
ATIVNSGFGPFSMASLPTFGAORPPSESSRARLWLTGNI GFGVNLVGLSAGSPR  
TDTAVADRLERLAFONLATHDDRLLSLILLARDHVHGWGLASGFMLCAAFNV  
LURGLCGRDTAAGPELVSASVRAVORLVAARDPVIIVILLAEGRRLDKLAVEA  
PEFHSANVLAELTLIGHRGPAEYMAATSVADNPPELLVRVAKTLRAVPAPQPTVIP  
LRKPAVLLAALQRLDRVRQRRLAIWLRALLKEYRGRUTEAGVFTDPDDVFLIL  
VDEILDPAVGLVARREQRRLAGIYPTVFGSEWEPSSSAALAAAGTDTLRGVG  
VCGGRVGRVTRVREPTIDDLQGEILVAEVDVGVYTAFCYAAAVTTEBLGPMHAA  
VVARFGPVCVDAQCATFLPFGALVEVDGATGEIHHVELASEDGPALPGSDLSR"  
complement (3332..15787)  
/gene="WT2108"  
complement (3332..15787)  
/gene="WT2108"  
/note="similar to PID:987099; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAK46387.1"  
/db\_xref="GI:13881781"  
/translation="WVDQLQHATEALRKALVQERLKRTRNALLERSSEPIAVGMSC  
RPFPGVDSPEGWQWADARDVNSEFTDRGWDLAFLDPDPDRHKSARTGGFVG

ORIGIN

Query Match

30.8%; Score 154.4; DB 1; Length 15970;

VADFDPAFFGISPSEALAMDPOHRMLLESWEALERAGIDPTGLRGSATGVPAGLIVG  
GYGLAEIEGYRLTGMTSSVASCRAVAVLGLGPAVSDVTACSSILVHMAVGSRLR  
SGECDLALAGVTYNATPTVFVEFSRHRGLAPDCECKPYAGRADGWGSEGGMLVLQ  
RLSDARRLGHVPLVAVVQASVNDQSGNCEGPEPLWGSVKNMHTTQAAGVSAEVDV  
VEGHTGTTTLTAEQALLATYQODRGEPEPLWGSVKNMHTTQAAGVSAEVDV  
VLAMRHELTPATLHVDPSPHVDWSAGAVELLTAPRVWMPAGARTAGVSGFISGNT  
AHVTIEAVVPRREAGWAGVPMWVSAGKSESALRQQAALAAVYRGGDGLDADVG  
WSLAGRSVEHRAVVGDRRLLAGLADGOLGGSVRGTAAGKTVFVFPQGG  
SQWLMGIELLDTAPFAQQIDACAEAFEDWSLVDVLRGAPCARLDVVDVQPV  
LFAVWVSLAEWLKSVAVHPDAVIGHSQGEIAAAVAGALSILDAARVTVILSKLAGL  
APGMSIACDAQDRLAPFGRVSVIAVWVSPSVAVVSGEVALGEBELIAEATKKE  
LTRRIEIVYASHSVEVEIARPLAELSGIEPSTRVTFVSTGNELDTAGLDADY  
WVRNVOTLVLPQAVRNACEQYRTFIESSPHAGITGVETFAACTDSDSIAVYPT  
LGRGDGLHRLLELLSAASAFVAGVANRGTLDGAGYVELPTVAFQKRFMLSAEGGA  
DYSGLGLSASHEPLIGAVVDLPASGVVLTGRLSPNVQPMWADHADVSLVLPFGTGV  
EALIRAGDEIVGCSVLDELTLAALPLPATGSAVVQVVDAGRDSNRGSVIFSRADAQ  
AGLILHAEGLIRPGSVEPGADLSVMPAGAVTVVADGYERLATIRGIRGPAFRGLITA  
MMARGEITFAEVLPEAAGVGGVGFHPALLDAHLVAVIAGDPDELALPFAMQGVSL  
HATGASAVARIAPAGPSAVSELADGLGLPVLVSVMVARPVTERRQLIAAVSGSQPL  
RLFEVIMPSAATSPGTPAVQIPESVAADQDPVAGSVVRSHQALAAVQSVLTHDES  
GVLVVTGAMALPREVDLAGAAGVGLVRSACQTEHPGRILVLDSDATDDAAATAMA  
LATGEPQVVLGGQVITARVGRSRAADAILVPGDGPWRLGLGSGAGTENLRLEPVPN  
ADAPLPGQGVYAMRAIAANFRDINITLGMFTDALLGEGGAGVVEVGPVTEFSVG  
DSVFGFPDGSSTLVAGDVRLLPMPADWSYAEAAAIASVFTYATYATHTLADVPQGG  
RVLIHAGTGGVGMAAVOLARHLGLEVFATASKGWDTLRMGFDHDDHISDSLSLEFED  
KFRANTGGRGPDVLDLSAGFEVDASLRVAPGGVLEMGKTDIRDGPVIAOQYPCGR  
YRAFLLPFGPRPMHOYMLELATLPGDGVLRPLVPTFDVPEAPALRYLSCARHTGK  
VYMLMPSGAAGTVLITGTGMAGSAVARHVVARHVRNMLVLSRRGPDAPGAELVA  
ELAAAGQVQVVAADRAALAKVADIPOVQHLGSIHTAGLADDAVMSLITPDVR  
DVLVASKYDAAMHLLHETLTDVSAFVMFSSMAGLVSGSQANYAANSFLDALAHR  
RAHGLPATISLGMGLWDQASMTGDLDAADLARLREGVLASTAEALFTDAMI VDE  
PELAPRIDLTLRAHAVAHPMFSDIASAPTRROVDDSVAAAKSKSALHRLHGLPE  
AEQHALVIGLRLHATATVILNITPEALDPKAFQDLGDESLTAVEMNELKSAATLSL  
PDDMMDLITQGRDSEFPDRGMDLAGLYNPDPAAGACYTRTGTFVDGVGDFDPAF  
FGVGESEALAMDPOQRMLLELSEALERAGIDPTGLRGSATGVPAGVMTQGYGMAAE  
PVEGRLTQGLSSVAGSVAVVYVLEGPASVSDTACSSILVHMAVGLSRGSECDLA  
LAGVTVNATPTVFVEFSRHRGLAPDCECKPYAGRADGWGSEGGMLVLQRLSDARR  
LGHFVLAVVGSVNDQASNLTPANGPSQQRVRAALANAGLSAEVDVVEGHTG  
TTLGDTIEAQLLATYQODRGEPEPLWGSVKNMHTTQAAGVSAEVDVVEGHTG  
LLPATLHVDPSPHVDWSAGAVELLTAPRVWMPAGARTAGVSGFISGNTAHVILIEA  
VVPVPRREAGWAGVPMWVSAGKSESALRQQAALAAVYRGGDGLDADVGWSLAGRS  
VPEHRAVVGDRRLLAGLADGOLGGSVRGTAAGKTVFVFPQGGSQWLMG  
MGLHAGYPVFAEAFNTVVGELDRHLLRLPLREVMWGHENLNTTEPQALFAVEVAL  
FELLSSKGVPRDPTFMGHSIGELSAARVAGVILENAVLVAARGLMQALPAGGAMVA  
VQAAEEVRPLLSAEVDIAVNGPSASLISGQNAVAVADQIRADGRVHQHQLASHA  
FHSPLDPMIDFEFAVAGIARTPTIGVISNVTQQLAGDGSARVRRHRLRQAVRF  
ADSVFAQAAGGSRFLEVPGSGLVASIEESLPDVAVTMTSALRDRPEPATLTNAVA  
QQFVTGMDLDMRAVVEGAOFVELPTVAFQRRFMLSQDGAADAAGLGLAASEHALLG  
AVIDL PASGVVLTGRLSPVQGLADHSVAGVTIFPGAGFVELA ITRAGDEVCQGVVD  
ELTLAAPLVLPASGSVAQVVTNGPDESQVRSVYSRGDTGCTGWLHARGALRAGSA  
EFTADLAWPFGAVFVEVADGTLQOLAEQYGYGPAFGLTAMWRGDEVPFAEVALPA  
DAGVTLQWPHFVILLDALHVAHVLASAEQSGSVVFFSWQVSLHAAAGSAVRA  
RIAPYSGAVSIELADGLGLPVLVSASMLAREVTDQQLRAAVSSSGPDRFLFVTVSPQ  
PSAAVEPLPCVCAAGTDESAAVVFESPLAGDVAGVAAATSSVLDLQSLTDRDAG  
VLVWTRGVALPGEEDVTDLAGAAVGLVRSACQTEHPGRIVLVDSADPLDLSALAAY  
TTGPEQVLRGDEVYTRVHGSRAVGGILLVPSDRPRLAMSTAGTFFENLRLELIPDA  
DTGLPGQVRVAVSAIAANFRDINITLGMFTDALLGEGGAGVVEVGPVTEFSVG  
DRVMLIFEGTGTVASTDPQRLVKVPAGWSTAAATTSVVFATAHYDLDAARSGQ  
RVLIHAGTGGVGMAAVOLARHLGLEVFATASKGWDTLRMGFDHDDHISDSLSLEFED  
KFRANTGGRGPDVLDLSAGFEVDASLRVAPGGVLEMGKTDIRDGPVIAOQYPCGR  
YRAFLLPFGPRPMHOYMLELATLPGDGVLRPLVPTFDVPEAPALRYLSOARHTGK  
VYMLMPSGAAGTVLITGTGMAGSAVARHVVARHVRNMLVLSRRGPDAPGAELVA  
ELAAAGQVQVVAADRAALAKVADIPOVQHLGSIHTAGLADDAVMSLITPDVR  
DVLVASKYDAAMHLLHETLTDVSAFVMFSSMAGLVSGSQANYAANSFLDALAHR  
RAHGLPATISLGMGLWDQASMTGDLDAADLARLREGVLASTAEALQI LFTAMI VDE  
PEMLPAHIDFAALKVKFGDGTUPMFWVDLINATPRROVDDSLAAAKSKSALLQRLLEGL  
PEDEQHAVLLDLVRSIATVLSASPEAIDPDRFQELGFDLSITAVEMNELKSAATLSL  
AUSPLTIFDYNPSAALAGTMRRELIGSSPDQTSVAAGAEALQRI VASIPVRLRQAG  
VLDDLLALANETETSGQDPALAPTAEQETADMDLDDLVNAAPRNODE"

Best Local Similarity 65.4%; Pred. No. 7.9e-08;  
 Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;

QY	5	TCCCGGAGCGGAGCAGACCGCATCCTGCTGAAGCTGGTCCGCGGCCACGCTTTCGACGG	64
Db	3825	TGCCCCGAAGACGAGCAACACGCGCTGCTGACCTGGTGGCTCGCACATCGGCCACCG	3766
QY	65	TGCTCGGCCACACGCGCGCGCGAAGCATCGGCGCCGCGCCAGGCGTTTCCAGGAGGTGGGT	124
Db	3765	TGCTGGGTAGCCAGCCCGCGAGCCATCGACCCGGATAGGGCATTCAGGAGTTGGGT	3706
QY	125	TGGAATCGTGGCGCGGCTAACCTCCGCAACAGCCTGCAACGCGGCCACCGGCTTCGCGC	184
Db	3705	TGGAATCGTCAACCGCGGTCGAAATCGCAACCGGCTCAAAATCGGCCACCGGCTGGGC	3646
QY	185	TGCCGCGGACGCTGATCTTGGACTACCCACCCCGGAGGCGCTGGTTCGGCTACCTGCGG	244
Db	3645	TTTCAACACGCTCATCTTGGACTACCCCACTCCGCGGCGCTGGCGGCTATATGCGTC	3586
QY	245	TGGAATCTCT-----GCGGAGGCGCGACGCGCTGGACGCGCGCGGAAGACGACC	295
Db	3585	GAGAACTGCTCGGCTCATCACCGCAAGACACTTCAGCGGTGGCGCGCGGGAAGCCGAC	3526
QY	296	TCCGCGGAGTCTCTCGCGGCGGTGCGGTTGCGCGGTTCAAGAGGCGGCGCTGCTGGACA	355
Db	3525	TACAAACGCAATTGTGGCGTCCATTCCGGTCAAGCGCTTACGGCAGGCGGAGTGTGATC	3466
QY	356	CGCTGCTCGGCTCGC	371
Db	3465	TGTGCTCGGCTGGC	3450

Search completed: June 17, 2004, 01:40:46  
 Job time : 2280.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 22:08:13 ; Search time 285 Seconds  
(without alignments)  
7467.885 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 ggggtcccgagcgagca.....cctcggcgagcagctga 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	501	100.0	53789	2 AAV21187	AAV21187 Amycolato
2	158.6	31.7	4689	3 AAZ87299	Aaz87299 S. venez
3	158.6	31.7	36778	3 AAZ87318	Aaz87318 S. venez
4	158.6	31.7	37948	3 AAZ87285	Aaz87285 S. venez
5	158.6	31.7	38506	3 AAAY5633	AAAY5633 Nucleotid
6	158.6	31.7	38506	3 AAZ56001	Aaz56001 Recombina
7	158.6	31.7	38506	3 ABE56090	ABE56090 S. venez
8	158.6	31.7	38506	7 ADA09418	Ada09418 Cosmid pk
9	155.8	31.1	4725	7 AAD55818	Aad55818 Micromono
10	155.8	31.1	60196	7 AAD55810	Aad55810 Micromono
11	154.4	30.8	110000	4 AA199682	AA199682 22
12	154.4	30.8	110000	4 AA199683	AA199683 22
13	152.6	30.5	65140	4 AAD17184	Continuation (23 o
14	152.6	30.5	125401	4 AAD17186	Aad17184 Streptom
15	148.8	29.7	11238	7 AAD55817	Aad55817 Micromono
16	144.8	28.9	113193	7 AAD55817	Aad55817 Streptom
17	142	28.3	27541	4 AAD17185	Aad17185 Streptom
18	141.2	28.2	2700	4 AAD17193	Aad17193 ERD48 ins
19	140.2	28.0	11220	3 AAZ87298	Aaz87298 S. venez
20	139.6	27.9	30690	3 AAAY92301	AAAY92301 S. avermi
21	139.6	27.9	30690	4 AAH79277	AAH79277 Streptom
22	135.8	27.1	43280	2 AAAT80413	AAAT80413 Ty lactone
23	134.2	26.8	3072	7 ACA37796	ACA37796 Prokaryot

24	133.6	26.7	5676	2 AAV21186	AAV21186 Amycolato
25	132.4	26.4	103599	4 ABX04971	ABX04971 S. cinnam
26	132	26.3	44377	2 AAT80414	Aat80414 Platenoli
27	132	26.3	44377	2 AAT78508	Aat78508 Platenoli
28	130.8	26.1	29879	2 AAQ46806	AAQ46806 eryA regi
29	130.8	26.1	103599	4 ABX04971	ABX04971 S. cinnam
30	127.4	25.4	50937	3 AAA09469	AAA09469 Streptoco
31	126.2	25.2	561	4 AAA89135	AAA89135 Polyketid
32	126	25.1	6360	7 ACC42735	Acc42735 Geldanamy
33	125.6	25.1	47981	4 AAF30757	Aaf30757 Micromono
34	124.8	24.9	1681	6 AAS18438	Aas18438 Contig 11
35	124.4	24.8	77536	3 AAA14651	AAA14651 Nucleotid
36	124	24.8	31422	3 AAA92302	AAA92302 S. avermi
37	124	24.8	31422	4 AAH79278	Aah79278 Streptomy
38	123	24.6	82746	7 AAL61224	Aal61224 Actinosyn
39	122	24.4	20394	4 AAF24892	Aaf24892 Pimaricin
40	122	24.4	77536	3 AAA14651	AAA14651 Nucleotid
41	120.6	24.1	75236	7 ABV75557	ABV75557 Saccharop
42	119.4	23.8	14055	7 AAL61170	Aal61170 Actinosyn
43	119.2	23.8	6210	7 ACC42734	Acc42734 Geldanamy
44	118.4	23.6	4466	3 AAA14663	AAA14663 Nucleotid
45	118.4	23.6	4478	3 AAA14661	AAA14661 Nucleotid

ALIGNMENTS

RESULT 1  
AAV21187  
ID AAV21187 standard; DNA; 53789 BP.  
XX  
AC AAV21187;  
XX  
DT 24-JUL-1998 (first entry)  
XX  
DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.  
XX  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin; ds.  
XX  
OS Amycolatopsis mediterranei.  
XX  
FH Key Location/Qualifiers  
CDS 1825..15543  
FT /\*tag= a  
FT /label= ORF\_A  
FT /product= "polyketide synthase"  
FT 15550..30759  
FT /\*tag= b  
FT /label= ORF\_B  
FT /product= "polyketide synthase"  
FT 30895..36060  
FT /\*tag= c  
FT /label= ORF\_C  
FT /product= "polyketide synthase"  
FT 36259..41325  
FT /\*tag= d  
FT /label= ORF\_D  
FT /product= "polyketide synthase"  
FT 41373..51614  
FT /\*tag= e  
FT /label= ORF\_E  
FT /product= "polyketide synthase"  
FT 51713..52993  
FT /\*tag= f  
FT /label= ORF\_F  
FT /product= "polyketide synthase"

WO9807868-A1.

26-FEB-1998.

18-AUG-1997; 97WO-EP004495.

XX PR 20-AUG-1996; 96BP-00810551.  
XX PA (NOVS ) NOVARTIS AG.  
XX PI Schupp T, Toupet C, Engel N;  
XX DR WPI; 1998-169172/15.  
XX DR P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX PT produce rifamycin and rifamycin analogues.  
XX PS Claim 4; Page 53-102; 205pp; English.  
XX CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX CC synthesis gene cluster DNA fragment from the present invention. The DNA  
XX CC fragment comprises a DNA region involved directly or indirectly in the  
XX CC gene cluster responsible for rifamycin synthesis, including the adjacent  
XX CC DNA regions to the right and left which, by reason of their function in  
XX CC connection with rifamycin biosynthesis, qualify as constituents of this  
XX CC rifamycin gene cluster, and functional fragments, derivatives or  
XX CC constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX CC gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX CC analogues or precursors. It can also be used for inactivating or  
XX CC modifying genes involved in anamycin or rifamycin biosynthesis. The DNA  
XX CC can be used for constructing mutant actinomycetes strains from which the  
XX CC natural rifamycin or anamycin biosynthesis gene cluster has been partly  
XX CC or completely deleted. The DNA fragment can be used for assembling a  
XX CC library of polyketide synthases, which can be used for assembling a  
XX CC library of polyketides. A hybridisation probe of the invention can be  
XX CC used for identifying DNA fragments involved in the biosynthesis of  
XX CC anamycins  
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 4.8e-71;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGGTCCCGCGCGAGCAGAACCGATCCGCTGAAAGCTGCTCGCGGGCCACGCTTCG 60  
Db 15043 GCGGTCCCGCGCGAGCAGAACCGATCCGCTGAAAGCTGCTCGCGGGCCACGCTTCG 15102  
QY 61 ACGGTGCTCGGCCACAGCGGCGCGAAGCGATCGCGCGCGCGAGCGCTTCAGAGGTC 120  
Db 15103 ACGGTGCTCGGCCACAGCGGCGCGAAGCGATCGCGCGCGCGAGCGCTTCAGAGGTC 15162  
QY 121 GCGTTCGACTCGCTGGCGCGCGTCAACCTCGCAACAGCTGCGCGGCGCACCGGGCTG 180  
Db 15163 GCGTTCGACTCGCTGGCGCGCGTCAACCTCGCAACAGCTGCGCGGCGCACCGGGCTG 15222  
QY 181 CGGCTGCCCGCGACGCTGATCTTCGACTACCCACCGGAGGCGCTGGTGGCTACCTG 240  
Db 15223 CGGCTGCCCGCGACGCTGATCTTCGACTACCCACCGGAGGCGCTGGTGGCTACCTG 15282  
QY 241 CGCGTCGAATCTCTCGGGAGGCGCGACGACGCGCTGAGCGGGGGAAGACACCTCCGG 300  
Db 15283 CGCGTCGAATCTCTCGGGAGGCGCGACGACGCGCTGAGCGGGGGAAGACACCTCCGG 15342  
QY 301 CGAGTCTCTCGGGCGCGTTCGCGCGGTTCAAGAGGCGGGCGTGTGACACGCTG 360  
Db 15343 CGAGTCTCTCGGGCGCGTTCGCGCGGTTCAAGAGGCGGGCGTGTGACACGCTG 15402  
QY 361 CTCGGCTCTCGCGACACCGGACCGGAACCGGACGCGCGCGGACACCGAGCGGCC 420  
Db 15403 CTCGGCTCTCGCGACACCGGACCGGAACCGGACGCGCGCGGACACCGAGCGGCC 15462  
QY 421 CGCGCGCGCGACGCGACGAACTGATCGACGCACTGGACATCTCCGGTCTCGTCAACGA 480  
Db 15463 CGCGCGCGCGACGCGACGAACTGATCGACGCACTGGACATCTCCGGTCTCGTCAACGA 15522  
QY 481 GCCCTCGGCGACGAGCTGA 501

Db 15523 GCCCTCGGCGACGAGCTGA 15543  
RESULT 2  
AAZ87299  
ID AAZ87299 standard; DNA; 4689 BP.  
XX AC AAZ87299;  
XX DT 15-SEP-2003 (revised)  
XX DT 05-JUN-2000 (first entry)  
XX DE S. venezuelae macrolide biosynthetic gene pikAIII, SEQ ID NO:34.  
XX KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
XX KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
XX KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
XX KW chronic obstructive pulmonary disease; respiratory inflammation;  
XX KW hypercholesterolaemia; crop protection agent; ds.  
XX OS Streptomyces venezuelae; ATCC15439.  
XX FH Key Location/Qualifiers  
XX FT CDS 1..4689  
XX FT /\*tag= a  
XX FT /product= "PikAIII"  
XX PN WO200000620-A2.  
XX PD 06-JAN-2000.  
XX PF 25-JUN-1999; 99WO-US014398.  
XX PR 26-JUN-1998; 98US-00105537.  
XX XX (MINU ) UNIV MINNESOTA.  
XX PI Sherman DH, Liu H, Xue Y, Zhao L;  
XX DR WPI; 2000-160679/14.  
XX DR P-PSDB; AAY77194.  
XX PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
XX PS synthesis of methymycin and pikromycin.  
XX Claim 15; Page 415-417; 438pp; English.  
XX CC The invention relates to an isolated and purified nucleic acid segment  
XX CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
XX CC biologically active variant, where the nucleic acid sequence is not  
XX CC derived from the eryC gene cluster of Saccharopolyspora erythraea or  
XX CC Streptomyces antibioticus. The invention also relates to a macrolide  
XX CC biosynthetic gene cluster, or fragments thereof. The macrolide  
XX CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
XX CC pikromycin, neomethymycin, narbomycin or a combination of these  
XX CC compounds. Recombinant or augmented cells comprising the desosamine  
XX CC and/or macrolide biosynthetic gene clusters are useful for the production  
XX CC of biologically active macrolides. The macrolide biosynthetic proteins  
XX CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
XX CC narbomycin. The alternative termination of polyketide synthase may be  
XX CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
XX CC monomers. The compounds produced by the recombinant host cells are useful  
XX CC as biopolymers, e.g., in packaging or biomedical applications, to  
XX CC engineer PHA monomer synthases or to prepare biologically active agents,  
XX CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
XX CC chronic obstructive pulmonary disease as well as other diseases involving  
XX CC respiratory inflammation, cholesterol-lowering agents or macrolide-based  
XX CC antibiotics which are active against a variety of organisms, e.g.,  
XX CC bacteria, including multi-drug resistant pneumococci and other  
XX CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
XX CC protection agents (e.g., fungicides or insecticides) via expression of  
XX CC polyketides in plants. Sequences AAZ87295-287302 represent macrolide

FT	CDS	26991..31679	
FT		/*tag= c	
FT		/product= "Pik gene cluster protein #3 (AAY77202)"	
FT	CDS	31782..35822	
FT		/*tag= d	
FT		/product= "pik gene cluster protein #4 (AAY77203)"	
FT	CDS	35819..36664	
FT		/*tag= e	
FT		/product= "Pik gene cluster protein #4 (AAY80997)"	
XX			
PN	W020000620-A2.		
XX			
PD	06-JAN-2000.		
XX			
PF	25-JUN-1999;	99WO-US014398.	
XX			
PR	26-JUN-1998;	98US-00105537.	
XX			
PA	(MINU ) UNIV MINNESOTA.		
XX			
PI	Sherman DH, Liu H, Xue Y, Zhao L;		
XX			
DR	WPI: 2000-160679/14		
DR	P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.		
XX			
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.		
PT	synthesis of methymycin and pikromycin.		
XX			
PS	Disclosure; Fig 31; 438pp; English.		
XX			
CC	The invention relates to an isolated and purified nucleic acid segment		
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its		
CC	biologically active variant, where the nucleic acid sequence is not		
CC	derived from the eryC gene cluster of Saccharopolyspora erythraea or		
CC	Streptomyces antibioticus. The invention also relates to a macrolide		
CC	biosynthetic gene cluster, or fragments thereof. The macrolide		
CC	biosynthetic gene cluster encodes proteins which synthesize methymycin,		
CC	pikromycin, neomethymycin, narbomycin or a combination of these		
CC	compounds. Recombinant or augmented cells comprising the desosamine		
CC	and/or macrolide biosynthetic gene clusters are useful for the production		
CC	of biologically active macrolides. The macrolide biosynthetic proteins		
CC	are useful for synthesis of methymycin, pikromycin, neomethymycin and		
CC	narbomycin. The alternative termination of polyketide synthesis may be		
CC	useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)		
CC	monomers. The compounds produced by the recombinant host cells are useful		
CC	as biopolymers, e.g., in packaging or biomedical applications, to		
CC	engineer PHA monomer syntheses or to prepare biologically active agents,		
CC	such as chemotherapeutics, immunosuppressants, agents to treat asthma,		
CC	chronic obstructive pulmonary disease as well as other diseases involving		
CC	respiratory inflammation, cholesterol-lowering agents or macrolide-based		
CC	antibiotics which are active against a variety of organisms, e.g.,		
CC	bacteria, including multi-drug resistant pneumococci and other		
CC	respiratory pathogens, as well as viral parasitic pathogens, or as crop		
CC	protection agents (e.g., fungicides or insecticides) via expression of		
CC	polyketides in plants. The present sequence represents the macrolide		
CC	biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439,		
CC	as given in figure 31. (Updated on 15-SEP-2003 to standardise OS field)		
XX			
XX	Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 0 U; 2 Other;		
SO			
	Query Match	31.7%;	Score 158.6; DB 3; Length 36778;
	Best Local Similarity	59.0%;	Pred. No. 5.6e-17;
	Matches	291; Conservative	0; Mismatches 199; Indels 3; Gaps 1;
QY	1	GCGGTCCCCGACGCCGAGAGAACCGCAPCTCTGCTGAAGCTGGTCCGGCGCCACCGCTCG	60
DB	31173	GCGCTCCCCGAGCCCGAGCGCGCGCGGCTCCTCACCTCTGTCGTACCCACGCGGGG	31239
QY	61	ACGGTCTGGCCACAGCGCGCGCGAGGCATCGGCCCGCCGCGCGGTTCAGAGGTC	120
DB	31233	GCGTACTCGGCCATTCCTCCCGCCGACCGGGGTGGCCCGCGCGCTGCTTCACCGAGTTC	31299
QY	121	GGCTTCGACTCGCTGGCGCGCGGTCAACCTTCGCAACAGCCTTCACGCGCTACCGGGCTG	180







```
PF XX 27-MAY-1999; 99WO-US011814.
PR XX 28-MAY-1998; 98US-0087080P.
PR XX 28-AUG-1998; 98US-00141908.
PR XX 22-SEP-1998; 98US-0108880P.
PR XX 08-FEB-1999; 99US-0119139P.
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX Ashley G, Betlach MC, Betlach M, Mcdaniel R, Tang L;
XX WPI; 2000-072618/06.
DR P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR AAY67208, AAY67211.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide synthase,
XX for production of ketolide antibiotics.
XX
PS Example 2; Page 16-27; 98pp; English.
XX
CC This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC Streptomyces venezuelae DNA insert. The cosmid contains open reading
CC frames which encode the various modules of the narbonolide polyketide
CC synthase (PKS). The invention relates to recombinant DNA containing a
CC coding sequence for a narbonolide PKS. Polyketides are compounds
CC synthesised from 2-carbon units through a series of condensations and
CC subsequent modifications. Modular PKSs are responsible for the production
CC of many antibiotics including picromycin. The narbonolide PKS consists of
CC a loading module, six extender modules, and two thioester domains. Four
CC proteins make up the narbonolide PKS (PICAI, PICAI, PICAI and PICAIIV).
CC PICAI includes the loading module and extender modules 1 and 2, PICAI
CC includes extender modules 3 and 4, PICAI includes extender module 5 and
CC second type II thioesterase domain is found on the PICB protein. The
CC nucleotide sequences encoding all of these proteins can be isolated in
CC recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC desaminylated in S. venezuelae to yield narbomycin, the desaminyl
CC transferase enzyme is required for this conversion, and the desaminyl
CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC DNA of the invention is used to express, in transformed cells,
CC narbonolide (or its derivatives) or other ketolides (particularly
CC hybrids), which may then be converted (e.g. by other enzymes
CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;

Query Match 31.7%; Score 158.6; DB 3; Length 38506;
Best Local Similarity 59.0%; Pred. No. 5.6e-17;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGACGCGAGCAGACGATCTCTGCTGAGCTGCTCGCGCGGCGACGCTCG 60
DB 29315 GCGTCTCCCGACCGAGCGCGCGCGCTCTCTACCTCGTCTCGTACCCACGCGGG 29374

QY 61 ACGGTGCTCGGCGACGCGCGCGGAGGATCGCGCGCGCGCGGCTTCAGGAGGTC 120
DB 29375 GCGGTACTCGGCGATTCCTCTCCCGGCGGCGCGCGCGGCTTCACCGAGCTC 29434

QY 121 GCGTTCGACTCGTGGCGCGGCTCACTTCGGAACAGCTTCGACGCGCGCGGCTG 180
DB 29435 GCGTTCGACTCGTTCGACGCGCGGCTGAGCTTCGCAACAGCTTCACGCGTGGTCCGAC 29494

QY 181 CGGCTGCGCGCGCGCTGATCTTCGATACCTACCCCGCGAGCGCTGCTCGCTACCTG 240
DB 29495 AGGCTCCCGCGCGCGCGCTTCGACCGCGCGCGCGCGCGCTTCGCGCGCGCTC 29554

QY 241 CGCGTTCGAACTCTCTCGGAGGCGCGCGCGCGCTTCGAGCGCGCGGAGACGCTCCGG 300
DB 29555 CACGAGCGCTACTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCTGGAGCGCGGTGCGC 29614

QY 301 CGAGTCTCTCGCGCGCGTCTCGTTCGCGCGCGTTCAGAGGCGCGCGTCTGTCGACACGCTG 360

Db 29615 CCGGCGCTCGCGCGACTGCCCTCGACCGGCTCGCGGCGCGGCTCTCGACCGCTC 29674
QY 361 CTCGCGCTCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
Db 29675 CTCGCGCTCACCGCGCATCGAGCCGAGCGCGGTTTCGCGCGGTTTCGAGCGCGCGCGCC 29734
QY 418 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477
Db 29735 GACCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29794
QY 478 CGAGCGCTCGCGCG 490
Db 29795 ATGGCTCTCGCGCG 29807

RESULT 7
ABS56090
ID ABS56090 standard; DNA; 38506 BP.
XX
AC ABS56090;
XX
DT 21-JAN-2003 (first entry)
XX
DE S. venezuelae DNA inserted into cosmid pKOS023-27.
XX
KW Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;
KW desosaminyl transferase gene; beta-glucosidase gene; antibiotic;
KW pick hydroxylase gene; C12 hydroxylase gene; narbonolide;
KW desosaminylated polyketide; narbomycin biosynthesis; mutant;
KW picromycin biosynthesis; da.
XX
OS Streptomyces venezuelae.
XX
SY Synthetic.
XX
PN WO200297062-A2.
XX
PD 05-DEC-2002.
XX
PF 22-FEB-2002; 2002WO-US005642.
XX
PR 22-FEB-2001; 2001US-00793708.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Mcdaniel R, Tang L;
XX WPI; 2003-041412/03.
XX
PT Preparation of polyketides by recombinant DNA technology, useful as
PT antibiotics and as intermediates in the synthesis of pharmaceutical
PT compounds.
XX
PS Disclosure; Page 20-30; 127pp; English.
XX
CC The present invention relates to recombinant DNA sequences encoding for a
CC narbonolide polyketide synthase (PKS) domain, and methods of producing
CC polyketides by recombinant DNA technology. The recombinant DNA sequences
CC are derived from Streptomyces venezuelae desosamine biosynthetic,
CC desosaminyl transferase, beta-glucosidase, or pick (C12) hydroxylase
CC genes. The method is useful for transforming a cell with a recombinant
CC expression vector that encodes a functional beta-glucosidase gene, and
CC therefore for increasing the yield of a desosaminylated polyketide in a
CC cell. The recombinant methods and materials are useful for expressing
CC polyketides with significant antibiotic activity, derived in whole or in
CC part from the narbonolide PKS gene, and other genes involved in
CC narbomycin and picromycin biosynthesis in recombinant host cells. The
CC present sequence represents S. venezuelae DNA inserted into cosmid
CC pKOS023-27 in the methods of the present invention
XX
SQ Sequence 38506 BP; 4914 A; 15116 C; 13446 G; 5030 T; 0 U; 0 Other;

Query Match 31.7%; Score 158.6; DB 7; Length 38506;
```

Best Local Similarity 59.0%; Pred. No. 5.6e-17;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGACCGCCGAGCGAAGCCGATCTCTGTAAGTGTTCGGGCGCAACGCTTCG 60  
DB 29315 GCGCTCCCGAGCGCCGAGCGCGGCGCTCTCTACCTCTGCTGATCCACGCGGCG 29374

QY 61 ACGGTGCTCGGCGACAGCGCGCGGAGGATCGCGCGCGCGCGCGCGCGCGCGCGGTC 120  
DB 29375 GCGGTACTCGGCGATCTCTCTCCCGGACCGGCGCGCGCGCGCGCGCGCGCGGTC 29434

QY 121 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGCTG 180  
DB 29435 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGTC 29494

QY 181 CGGCTCGCGCGAGCGCTGATCTTCGACTACCCCGCGAGCGCGCGCGCGCGCGCGGTC 240  
DB 29495 AGGCTCCCGGTC 29554

QY 241 CGGCTCGGCGGTC 300  
DB 29555 AGGCTCCCGGTC 300

QY 301 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGCTG 360  
DB 29615 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGTC 360

QY 361 CTGCGGCTCGGTC 417  
DB 29675 CTGCGGCTCGGTC 417

QY 418 GCGGTC 477  
DB 29735 GCGGTC 477

QY 478 CGAGCGGTC 490  
DB 29795 ATGGCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 490

RESULT 8  
ADA09418  
ID ADA09418 standard; DNA; 38506 BP.  
AC ADA09418;  
XT 06-NOV-2003 (first entry)  
XX Cosmid pKOS023-27 containing S. venezuelae PKS gene cluster.  
XX Streptomyces venezuelae; Sv; narbonolide polyketide synthase; PKS;  
KW narbonolide PKS; narbomycin biosynthesis; picromycin biosynthesis;  
KW PKS gene cluster; picromycin; glycosylation; hydroxylation;  
KW Cl2 hydroxylase; Pick; desosamine biosynthesis;  
KW desosaminyl transferase enzyme; antibiotic; narbonolide synthase; gene;  
ds.  
XX Synthetic.  
OS Streptomyces venezuelae.  
XX US6509455-B1.  
XX 21-JAN-2003.  
XX 07-SEP-2000; 2000US-00657440.  
XX 30-APR-1997; 97US-00846247.  
XX 06-MAY-1998; 98US-00073538.  
XX 28-MAY-1998; 98US-0087080P.  
XX 28-AUG-1998; 98US-00141908.  
XX 22-SEP-1998; 98US-0100880P.  
XX 08-FEB-1999; 99US-0119139P.  
XX 20-MAY-1999; 99US-0134990P.

PR 27-MAY-1999; 99US-00320878.  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
XX Ashley G, Betlach MC, Betlach M, Modaniel R, Tang L;  
XX WPI; 2003-352291/33.  
XX Novel recombinant DNA compounds comprising coding sequences for  
PT desosamine transferase gene of Streptomyces venezuelae, useful for  
PT producing desosamine transferase which transfers desosamine to substrate  
XX polyketides.  
XX Claim 2; Col 21-54; 132pp; English.  
XX The present invention relates to recombinant DNA compounds that encode  
CC Streptomyces venezuelae (Sv) narbonolide polyketide synthases (PKSs). The  
CC recombinant PKSs are derived from narbonolide PKS and other genes  
CC involved in narbomycin and picromycin biosynthesis in recombinant host  
CC cells. The invention also discloses the S. venezuelae PKS gene cluster  
CC that results in the production of picromycin. Also disclosed are enzymes  
CC such as those responsible for glycosylation and hydroxylation, (e.g. Cl2  
CC hydroxylase (Pick)), desosamine biosynthesis, and desosaminyl transferase  
CC enzymes. The recombinant narbonolide, narbonolide derivatives, and  
CC polyketides are useful as antibiotics and as intermediates in the  
CC synthesis of compounds for pharmaceutical applications. The present  
CC sequence represents cosmid pKOS023-27 containing S. venezuelae PKS gene  
CC cluster.  
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 U; 0 Other;  
Query Match 31.7%; Score 158.6; DB 7; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 5.6e-17;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGACCGCCGAGCGAAGCCGATCTCTGTAAGTGTTCGGGCGCAACGCTTCG 60  
DB 29315 GCGGTCCCGAGCGCCGAGCGCGGCGCTCTCTACCTCTGCTGATCCACGCGGCG 29374

QY 61 ACGGTGCTCGGCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 120  
DB 29375 GCGGTACTCGGCGATCTCTCTCCCGGACCGGCGCGCGCGCGCGCGCGGTC 29434

QY 121 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGCTG 180  
DB 29435 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGGTC 29494

QY 181 CGGCTCGCGCGAGCGCTGATCTTCGACTACCCCGCGAGCGCGCGCGCGCGCGGTC 240  
DB 29495 AGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 29554

QY 241 CGGCTCGGCGGTC 300  
DB 29555 AGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 300

QY 301 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGCGGCTG 360  
DB 29615 GCGTTCGACTCGTGGCGCGGCTCAACCTCGGCAACAGCCTGACGCGCGCGCGGTC 360

QY 361 CTGCGGCTCGGTC 417  
DB 29675 CTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 417

QY 418 GCGGTC 477  
DB 29735 GCGGTC 477

QY 478 CGAGCGGTC 490  
DB 29795 ATGGCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 490

RESULT 9

```

AAD55818
ID AAD55818 standard; DNA; 4725 BP.
AC
AC AAD55818;
XX
XX
XX
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
DE
DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #8.
XX
XX Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW Gene; ds.
XX
XX Micromonospora carbonacea.
OS
OS
OS Key Location/Qualifiers
FH 1..4725
FT CDS
FT /*tag= a
FT /product= "Polyketide synthase"
XX
XX CA2391131-A1.
PN
PN 19-NOV-2002.
PD
PD 26-JUL-2002; 2002CA-02391131.
PF
PF 26-JUL-2001; 2001US-0307629P.
PR
PR (ECOP-) ECOPIA BIOSCIENCES INC.
PA
PA Yang X, Staffa A, Farnet CM;
PI WPI; 2003-343556/33.
XX DR P-PSDB; AAE37002.
XX
XX Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of rosaramicin.
XX
XX Claim 1; Page 164-166; 206pp; English.
XX
XX The invention relates to genes and proteins involved in the biosynthesis
CC of macrolides by microorganisms. In particular it relates to the nucleic
CC acids forming the biosynthetic locus for rosaramicin (a 16-member
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
CC useful for the biosynthesis of macrolides by microorganisms. It allows
CC direct manipulation of macrolides and related chemical structures by
CC chemical engineering of the proteins involved in the biosynthesis of
CC rosaramicin. It is useful to catalyze certain biochemical reactions, in
CC vitro or in vivo, to direct or enhance the synthesis or modification of a
CC polyketide, polyketide substrate or its precursor. The present sequence
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
CC OCT-2003 to standardise OS field)
XX
XX Sequence 4725 BP; 482 A; 1782 C; 1885 G; 576 T; 0 U; 0 Other;
SQ
Query Match 31.1%; Score 155.8; DB 7; Length 4725;
Best Local Similarity 62.2%; Pred. No. 2.1e-16;
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;
QY 12 CCGCGAGCAGACCGGATCTGCTGAGCTGTGTCGCGCGCCACGCTTCGACCGGTGCTCGG 71
DB 4254 CGACGAACCGCGCGGGGCTGTGCTGACCTGTGATCGCGCGAGTCCGCGCGTCTCGG 4313
QY 72 CCACAGCGCGCGGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
DB 4314 CCACCGCGCGCGCGGAGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4373
QY 132 GCTGGCGCGCGGTCAACCTCCGCAACAGCTGTGACGCGCGCGCGCGCGCGCGCGCGCG 191
DB 4374 GTGACCGCGCGGTGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4433
QY 192 GACGCTGATCTTTCGATACCCACCCCGGAGGCGGTGTCGCGCTACCTCGCGGTGGAAT 251

```

---

```

Db 4434 CACCTCTGCTTTCGACCACTCCCGGACGCGGCTCGCGGACCTCGCGGCTGCT 4493
QY 252 CCTGCGGAGCGCGACGACGCGCTTGAGCGGCGGGAAGACGACCTTCGCGGAGTCTCGC 311
Db 4494 CGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4550
QY 312 GCGCGTCCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
Db 4551 CGACCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4610
QY 372 CGACACCGGACCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431
Db 4611 GGGGCTGGAGCGCGACGCGGTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 4670
QY 432 CGA 434
Db 4671 CGA 4673

```

---

```

RESULT 10
AAD55810
ID AAD55810 standard; DNA; 60196 BP.
XX
XX AAD55810;
XX
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
XX Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.
DE
DE Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW gene; ds.
XX
XX Micromonospora carbonacea.
OS
OS Key Location/Qualifiers
FH 1..1683
FT CDS
FT /*tag= a
FT /product= "Polyketide synthase #1"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (1728..2522)
FT /*tag= b
FT /product= "Polyketide synthase #2"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (2629..3861)
FT /*tag= c
FT /product= "Polyketide synthase #3"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (3965..5573)
FT /*tag= d
FT /product= "Polyketide synthase #4"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (5702..19117)
FT /*tag= e
FT /product= "Polyketide synthase #5"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (19144..24921)
FT /*tag= f
FT /product= "Polyketide synthase #6"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (24993..36230)
FT /*tag= g
FT /product= "Polyketide synthase #7"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (36292..41016)
FT /*tag= h
FT /product= "Polyketide synthase #8"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (41049..46403)
FT /*tag= i
FT /product= "Polyketide synthase #9"
FT /note= "CDS does not include start codon"
FT /partial
FT complement (46400..47794)
FT /*tag= j
FT /product= "Polyketide synthase #10"
FT /note= "CDS does not include start codon"
FT /partial

```

```
FT CDS 47794..49083
FT     /*tag= k
FT /product= "Polyketide synthase #11"
FT 49092..49814
FT     /*tag= l
FT /product= "Polyketide synthase #12"
FT 49868..51226
FT     /*tag= m
FT /product= "Polyketide synthase #13"
FT /transl_except= (pos:51221..51223, aa:Xaa)
FT /note= "Xaa corresponds to amino acids from position 452-1811"
FT CDS 51506..53416
FT     /*tag= n
FT /product= "Polyketide synthase #14"
FT /note= "CDS does not include start codon"
FT /partial
FT complement(54569..53358)
FT CDS
FT /product= "Polyketide synthase #15"
FT 54997..56342
FT     /*tag= p
FT /product= "Polyketide synthase #16"
FT 56408..57634
FT     /*tag= q
FT /product= "Polyketide synthase #17"
FT 57657..59123
FT     /*tag= r
FT /product= "Polyketide synthase #18"
FT /note= "CDS does not include start codon"
FT /partial
FT 59363..60196
FT     /*tag= s
FT /product= "Polyketide synthase #19"
FT CDS
FT CA2391131-A1.
FT 19-NOV-2002.
FT 26-JUL-2002; 2002CA-02391131.
FT XX
FT 26-JUL-2001; 2001US-0307629P.
FT PR (ECOP-) ECOPIA BIOSCIENCES INC.
FT PA Yang X, Staffa A, Farnet CM;
FT PI
FT XX
FT DR WPT; 2003-343556/33.
FT DR P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,
FT DR AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,
FT DR AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.
FT XX
FT PT Novel isolated polypeptide involved in biosynthesis of macrolides by
FT microorganisms, useful for biosynthesis of macrolides by microorganisms,
FT preferably for biosynthesis of rosamycin.
FT FT
FT PS Example 2; Page 59-94; 206pp; English.
FT XX
FT CC The invention relates to genes and proteins involved in the biosynthesis
FT of macrolides by microorganisms. In particular it relates to the nucleic
FT acids forming the biosynthetic locus for rosamycin (a 16-member
FT macrolide antibiotic) from Micromonospora carbonacea. The invention is
FT useful for the biosynthesis of macrolides by microorganisms. It allows
FT direct manipulation of macrolides and related chemical structures by
FT chemical engineering of the proteins involved in the biosynthesis of
FT rosamycin. It is useful to catalyse certain biochemical reactions, in
FT vitro or in vivo, to direct or enhance the synthesis or modification of a
FT polyketide, polyketide substrate or its precursor. The present sequence
FT is M. carbonacea polyketide synthase (PKS) type I gene cluster. (Updated
FT on 27-OCT-2003 to standardise OS field)
FT CC
FT XX Sequence 60196 BP; 6934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;
FT SQ
```

```
Query Match 31.1%; Score 155.8; DB 7; Length 60196;
Best Local Similarity 62.2%; Pred. No. 1.4e-16;
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

Qy 12 GCGGAGCAGAAACCGCATCTCTGCTGAAGCTGCTGCGGGCCACGCTTCGACGGTGTCTCGG 71
Db 40545 CGACGAAACGCGCGCGGGGCTGCTCGACCTGGTACGCGGCGAGGTCGCGCGCTCTCGG 40604

Qy 72 CCACAGCGGCGCGCGAAGGCGATCGGCGCGCGCGAGCGGCTTCAGAGAGGTTCGACTC 131
Db 40605 CCACCGCGGCGCGCGAGCAGCTCGGCGCGCGCGAGCGCGGCTTCGCGAGATCGGATTGACTC 40664

Qy 132 GCTGGCGCGCGGTCAACCTCGCGCAACAGCTGCGCGCGCGCGCGCGCGCGCTGCGGCT 191
Db 40665 GCTGACCGCGCTCGACCTGGCGCAAGCGGCTCAGGCGCGCGGTCGCGGCTGTCTCGC 40724

Qy 192 GAGCTGATCTTCGACTACCCCGGAGGCGCTGGTGGGCTACCTGCGGCTCGACT 251
Db 40725 CACCTCGTCTTCGACCAACCCCGCGCGCGCGCGCGCGCGCGCTGCGGCTGCT 40784

Qy 252 CTTGCGGCGGCGCGACGACGCGCTGCGAGCGGCGGCGGAGACGACCTCGGCGGAGTCTCGC 311
Db 40785 CGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40841

Qy 312 GCGCGTCCGCTTCGCGCGGTTCAAGAGGCGGCGGCTGTGACACGCTGCTCGGCTCGC 371
Db 40842 CGACCTGCGCTGCGCGCGGCTGCGGCGCGCGCGCGCGCGCGCTTCTGCGCTTGC 40901

Qy 372 CGACACCGGCGCGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431
Db 40902 GGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431

Qy 432 CGA 434
Db 40962 CGA 40964

RESULT 11
AAI99682 22/c
Continuation (23 of 45) of AAI99682 from base 2200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000
WP AAI99682_09 900001 1010000
WP AAI99682_10 1000001 1110000
WP AAI99682_11 1100001 1210000
WP AAI99682_12 1200001 1310000
WP AAI99682_13 1300001 1410000
WP AAI99682_14 1400001 1510000
WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
```

WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 30.8%; Score 154.4; DB 4; Length 110000;  
 Best Local Similarity 65.4%; Pred. No. 2.2e-16;  
 Matches 246; Conservative 0; Mismatches 121; Indels 9; Gaps 1;

QY	5	TCCCGGACGCGGAGCAGAACCGCATCTGCTGAAGCTGTGTCGCGGCACGCTTCGACGG	64
Db	95022	TGCCCGAAGACGAGCAACACGCGCTGCTGCTGGACCTGGTGGCTCGCACATCGCCACCG	94963
QY	65	TGCTCGGCCACAGGGGCGCGGAGGCATCGGCCGCGGCAGGCGTTCAGGAGTGGCT	124
Db	94962	TGCTGGGTAGCGGCAGCGCCGGAAGCCATCAGCCGGATAGGGCATTCAGAGAGTTGGGTT	94903
QY	125	TGCACTCGCTGGCGCGCGGTCAACTCCGCAACAGCCTGCACGCGGCACCGGGCTCGGC	184
Db	94902	TGCACTCGCTACCGCGGTGCAATATGGGCAACGCGCTCAATCGCACCGGCTGGCGC	94843
QY	185	TGCCCGCGAGCGCTGATCTTGCACTACCCACCCCGGAGGCGCTGCTGGGTACCTCGCGC	244
Db	94842	TTTCAACCCAGCTCATCTTCGACTACCCAACTCCGCGGCGTGGCGGCTATATCGGTC	94783
QY	245	TGCAACTCCT-----CGGGAGGCGCAGACGCGCTTGAGCGGCGGGAAGACGACC	295
Db	94782	GAGAACTGCTTCGGCTCATCAACCGCAGACACTTCAGCCGTGGCGGCGGGGAAAGCGAAC	94723
QY	296	TCCGGGAGTCCCTCGCGGCGCTCGCTTCGCCCGGTTCAAGAGAGCGGGCGTCTGGACA	355
Db	94722	TACAAGCATTTGGTGGGTTCATTCCGGTCAAGCGCTTACGGCAGCGGGAGTCTGGATC	94663
QY	356	CGCTGTTCGGCCTCGC	371
Db	94662	TGTTGCTCGGCTGGC	94647

RESULT 12  
 AAI99683\_22/c  
 Continuation (23 of 44) of AAI99683 from base 2200001 (Mycobacterium tuberculosis strain  
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683  
 WP Fragment Name Begin End  
 WP AAI99683\_00 1 110000  
 WP AAI99683\_01 100001 210000  
 WP AAI99683\_02 200001 310000  
 WP AAI99683\_03 300001 410000  
 WP AAI99683\_04 400001 510000  
 WP AAI99683\_05 500001 610000  
 WP AAI99683\_06 600001 710000  
 WP AAI99683\_07 700001 810000  
 WP AAI99683\_08 800001 910000  
 WP AAI99683\_09 900001 1010000  
 WP AAI99683\_10 1000001 1110000  
 WP AAI99683\_11 1100001 1210000  
 WP AAI99683\_12 1200001 1310000  
 WP AAI99683\_13 1300001 1410000  
 WP AAI99683\_14 1400001 1510000  
 WP AAI99683\_15 1500001 1610000  
 WP AAI99683\_16 1600001 1710000  
 WP AAI99683\_17 1700001 1810000  
 WP AAI99683\_18 1800001 1910000  
 WP AAI99683\_19 1900001 2010000



```

FT CDS /product= "NysA protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement(58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement(59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement(60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"
FT
FT WO200159126-A2.
FT
FT 16-AUG-2001.
FT
FT 08-FEB-2001; 2001WO-GB000509.
FT
FT 08-FEB-2000; 2000GB-00002840.
FT 10-APR-2000; 2000GB-00008786.
FT 14-APR-2000; 2000GB-00009387.
FT
FT (UNNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
FT (SNF) SINTEF STIFTELSEN IND TEK FORSK.
FT (ALPH-) ALPHARMA AS.
FT (SINV-) SINVENT AS.
FT (DZIE-) DZIELEWSKA H.
FT (ZOTC-) ZOTCHEV S B.
FT (SEKU-) SEKUROVA O N.
FT (FJAE-) FJAEVRIK E.
FT (BRAU-) BRAUTASET T.
FT (STRO-) STROM A R.
FT (VALL-) VALLA S.
FT
FT Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
FT Valla S, Ellingsen TE, Sletta H, Gulliksen O;
FT
FT WPI; 2001-557614/62.
FT P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
FT AAE10149, AAE10150.
FT
FT New nystatin polyketide synthase polynucleotides and polypeptides, useful
FT as antibiotics and antifungals.
FT
FT Claim 1; Page 188-254; 266pp; English.
FT
FT The present invention relates to the cloning and sequencing of the gene
FT cluster encoding a modular type I polyketide synthase (PKS) enzyme
FT involved in the biosynthesis of the macrolide antibiotic nystatin. The
FT nystatin PKS is useful as antifungal antibiotics. The present sequence is
FT a Streptomyces noursei nystatin PKS gene cluster DNA
FT
FT Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;
FT
FT Query Match 30.5%; Score 152.6; DB 4; Length 125401;
FT Best Local Similarity 60.7%; Pred. No. 4.2e-16;
FT Matches 276; Conservative 0; Mismatches 164; Indels 15; Gaps 1;
FT
FT 1 GCGGTCCCGCCGACGACGACCGCATCTGCTGAGCTGTCGCGCGCACGCTTCG 60
FT 76291 GAGCTGTCCGCGCGGCGCGCGCGCTGGTCGAGCGCTCCGCGCGGAGGTC 60
FT
FT 61 ACGGTGCTCGCCACAGCGCGCGCGGAGGATCGGCGCGCGGCTTCCAGAGGTC 120

```

```

Db 76351 GCGACCTCGGCCACGACACCCCGACGCGCATCCCGCGCGCTTCCGCGAGGTC 76410
Qy 121 GGTTCGACTCGTGGCGCGGTCAACCTCGGCAACGCTGCAACGCGGCGCCACCGGGCTG 180
Db 76411 GGTTCGACTCGTGGCGCGGTCAACCTCGGCAACGCTGCAACGCGGCTTCCGCGAGGTC 76470
Qy 181 CGGCTGCGCGGACGCTGATCTTCGACTACCCACCCCGGAGGCGCTGGTTCGGTACCTG 240
Db 76471 CGGCTGCGCGGCGCTGCTTCGACTACCCACCCCGGAGGCGCTGGTTCGGTACCTG 76530
Qy 241 CGGCTGCAACTCTTCGCGGAGGCGCGAGA-----CGGCTGCGAGCGGGGG 285
Db 76531 GCGCGCTGCTCTTCGCGCACCGCGCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCG 76590
Qy 286 GAAGACGACCTCGCGCGAGTCTCGCGGCGCTGCGCTTCGCGCGCGCGCGCGCGCGCGCG 345
Db 76591 GACGCCGCGCATCCGCGAGGCGCTCGCACCGCTCCCGATCGGACGCGCTCGGCAAGCGGGC 76650
Qy 346 GTGCTGACACGCTGCTCGGCTCGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
Db 76651 CTCCTGACATGCTGTAACCTCGCGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 76710
Qy 406 ACCACCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
Db 76711 GCCGACGCCCGCTCGGAATCGCTCGACGACATGGA 76745

RESULT 15
AAD55817
ID AAD55817 standard; DNA; 11238 BP.
XX
AC AAD55817;
XX
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DE Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.
XX
KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;
KW gene; ds.
XX
OS Micromonospora carbonacea.
XX
FH Key Location/Qualifiers
FT CDS 1..11238
FT /*tag= a
FT /product= "polyketide synthase"
FT /note= "CDS does not include start codon"
FT /partial
XX
CA2391131-Al.
XX
XX 19-NOV-2002.
PD
XX
XX 26-JUL-2002; 2002CA-02391131.
PF
XX
XX 26-JUL-2001; 2001US-0307629P.
PR
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
PA
XX
XX Yang X, Staffa A, Farnet CM;
PI
XX
XX WPI; 2003-343556/33.
DR
XX
XX P-PSDB; AAE37001.
DR
PT Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of rosamycin.
XX
XX Claim 1; Page 151-158; 206pp; English.
PS
XX
XX The invention relates to genes and proteins involved in the biosynthesis
CC

```

CC of macrolides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for roaramycin (a 16-member  
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macrolides by microorganisms. It allows  
CC direct manipulation of macrolides and related chemical structures by  
CC chemical engineering of the proteins involved in the biosynthesis of  
CC roaramycin. It is useful to catalyze certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polyketide, polyketide substrate or its precursor. The present sequence  
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-  
CC OCT-2003 to standardise OS field)  
XX  
SQ Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;  
Query Match 29.7%; Score 148.8; DB 7; Length 11238;  
Best Local Similarity 58.6%; Pred. No. 2.4e-15;  
Matches 258; Conservative 0; Mismatches 182; Indels 0; Gaps 0;  
QY 1 GCGGTCCCGGACGCGGAGCAGAACCGCATCTCTGCTGAAGCTGGTCGGGGCCACCGCTTCG 60  
Db 10717 GGGCTGACCGTGGCCGGAACAGCGGGAGCTGGTCTGGAGGCGTGGCGGACACGCGGCG 10776  
QY 61 ACGGTGCTCGGCACAGCGCGCGCGAAGCATTCGGCCCGCGCCAGGCGTTCCAGAGAGTTC 120  
Db 10777 GCGGTCTCTCGGACACGCCGACCGCGAGCGCGTCCGACCGCGGCGCTTCGGGAACTC 10836  
QY 121 GGCCTTCGACTCGGTGGCCCGCGGTCAACCTCCGCAACAGAGCTGACCGCGGCCACCGGGCTG 180  
Db 10837 GGCCTTCGACTCGGTGACGCGCGGTGGAGCTGGCAATCGGCTGGCCACCGCGTCCGGGCTG 10896  
QY 181 CGGCTGCCCGGAGCGCTGATCTTCGACTACCCACCCCGGAGGCGCTGGTGGGTACCTG 240  
Db 10897 CGCCTGCCCGGAGCGCTGCTTCGACCAACCCCGGAGCGCTGGCGGAGCACCTG 10956  
QY 241 CGGCTCGAACTCTCGCGGAGGCGGACGCGCTGGACGGCGGGAAGACGACCTCCGG 300  
Db 10957 CTCGCGGGGCTCGCGCCCGGAGCGGCGCGGCGGCGAGTTGCCGTTGTCGCCGAGCTGGGC 11016  
QY 301 CGAGTCTCTCGCGGCGCTGCGCTTCGCGCCCGGTTCAAGGAGGCGGCGCTGTGTHACAGCTG 360  
Db 11017 CGGCTGGAGGCGGCGCTGGCGCGCCACCGAGCGGCGCGCCCTCGAGCGGCTGGAGCACCTG 11076  
QY 361 CTCGGGCTTCGCGACACCGGACCGGAACCGGGGACCGGACCGGAGACCGGAGCGGCC 420  
Db 11077 GTGCGCGGAGGTGGGCTCGGATCGCGCGCTGGCGCGCTGGCGCGCTGGCGCGCGGC 11136  
QY 421 CCGGCGCGCGGACGCGCAGA 440  
Db 11137 GACGACGTGGCGGACCGA 11156

Search completed: June 16, 2004, 23:46:44  
Job time : 288 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:22:28 ; Search time 54.6667 Seconds  
(without alignment)  
5085.922 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggtccccgacgcgagca.....cctcgggcagacagctga 501

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158.6	31.7	4689	3 US-09-105-537-34	Sequence 34, Appli
2	158.6	31.7	36778	3 US-09-105-537-5	Sequence 5, Appli
3	158.6	31.7	38506	3 US-09-320-878-19	Sequence 19, Appli
4	158.6	31.7	38506	4 US-09-141-908-1	Sequence 1, Appli
5	158.6	31.7	38506	4 US-09-657-440-19	Sequence 19, Appli
C 6	154.4	30.8	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 7	154.4	30.8	4411529	3 US-09-103-840A-1	Sequence 1, Appli
8	140.2	28.0	11220	3 US-09-105-537-32	Sequence 12, Appli
9	135.8	27.1	43280	2 US-08-804-227C-1	Sequence 1, Appli
10	132	26.3	44377	2 US-08-804-227C-7	Sequence 7, Appli
11	132	26.3	44377	2 US-08-804-198-1	Sequence 1, Appli
12	130.8	26.1	11219	1 US-07-642-734C-1	Sequence 1, Appli
13	130.8	26.1	11219	3 US-08-439-009A-1	Sequence 1, Appli
14	127.4	25.4	50937	3 US-09-428-517-1	Sequence 1, Appli
C 15	126.2	25.2	561	3 US-09-154-083-13	Sequence 13, Appli
16	125.6	25.1	47981	4 US-09-679-279-1	Sequence 1, Appli
17	124.8	24.9	1681	4 US-09-434-288-7	Sequence 7, Appli
18	124.4	24.8	77536	4 US-09-410-551B-1	Sequence 1, Appli
C 19	122	24.4	77536	4 US-09-410-551B-1	Sequence 1, Appli
20	118.4	23.6	4466	4 US-09-410-551B-20	Sequence 20, Appli
21	118.4	23.6	4478	4 US-09-410-551B-16	Sequence 16, Appli
22	118.4	23.6	4547	4 US-09-410-551B-22	Sequence 22, Appli
23	118.4	23.6	4571	4 US-09-410-551B-18	Sequence 18, Appli
24	117.2	23.4	80161	3 US-09-036-987A-1	Sequence 1, Appli
25	117.2	23.4	80161	3 US-09-370-700-1	Sequence 1, Appli
26	117.2	23.4	80161	4 US-09-603-207-1	Sequence 1, Appli
27	117	23.4	20235	1 US-07-642-734C-3	Sequence 3, Appli

28	117	23.4	20235	3 US-08-439-009A-3	Sequence 3, Appli
29	115.4	23.0	13842	3 US-09-105-537-30	Sequence 30, Appli
C 30	110.2	22.0	1434	4 US-09-434-288-3	Sequence 3, Appli
31	108.4	21.6	3978	4 US-09-266-965-19	Sequence 19, Appli
32	108.4	21.6	12249	4 US-09-266-965-74	Sequence 74, Appli
33	108.4	21.6	18331	4 US-09-266-965-96	Sequence 96, Appli
34	107.4	21.4	15872	3 US-09-105-537-1	Sequence 1, Appli
35	107.4	21.4	15872	4 US-09-091-609-1	Sequence 1, Appli
36	107.4	21.4	15872	4 US-09-091-609-3	Sequence 3, Appli
37	104	20.8	990	4 US-09-266-965-18	Sequence 18, Appli
C 38	102	20.4	748	3 US-09-154-083-1	Sequence 1, Appli
C 39	102	20.4	753	3 US-09-154-083-28	Sequence 28, Appli
40	100.6	20.1	13987	2 US-08-804-227C-13	Sequence 13, Appli
41	98.6	19.7	28958	1 US-08-258-261B-6	Sequence 6, Appli
42	98.6	19.7	28958	1 US-08-456-837-6	Sequence 6, Appli
43	98.6	19.7	28958	1 US-08-457-342-6	Sequence 6, Appli
44	98.6	19.7	28958	1 US-08-457-646A-6	Sequence 6, Appli
45	98.6	19.7	28958	1 US-08-458-076A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-105-537-34  
; Sequence 34, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; TYPE: DNA  
; LENGTH: 4689  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-34

Query Match	31.7%;	Score	158.6;	DB	3;	Length	4689;
Best Local Similarity	59.0%;	Pred. No.	1.4e-20;				
Matches	291;	Conservative	0;	Mismatches	199;	Indels	3;
						Gaps	1;
QY	1	GC	GCCTCCCGACGCCGAGCAGAACCGCATCTCTGAGCTGCTCCGCGGCCACGCTTCG	60			
Db	4183	GC	GCCTCCCGACGCCGAGCAGCGCGGCGCTCTCACCTCTGCTCCATCCACCGCGCG	4242			
QY	61	AC	GCCTCCCGACGCCGAGCAGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	120			
Db	4243	GC	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4302			
QY	121	GC	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	180			
Db	4303	GC	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4362			
QY	181	CG	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	240			
Db	4363	AG	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4422			
QY	241	CG	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	300			
Db	4423	CAC	GAGCGGTACCTCGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4482			
QY	301	CG	AGTCTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	360			
Db	4483	CG	GCCTCCCGACGCCGAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4542			
QY	361	CT	CGCGCTCGCGCACCGGCACCGGCACCGGCACCGGCACCGGCACCGGCACCGGCAC	417			

```

db      4543  CTGGCGCCTCACGGGATCGAGCCCGAGCCGGGTTCGGCGGCGCGCCGCC 4602
QY      418  GCCCGCGCGCGACGAGCGAGAACTGATGACGCACTGGACATCTCCGGTCTCGTGCAA 477
db      4603  GACCTCTGGTGGCGAGCGCGAGCGTCGATGACGACCTGGAGCGCGAGGCCCTGATCGG 4662
QY      478  CGAGCCCTCGGCG 490
db      4663  ATGGCTCTCGGCC 4675

RESULT 2
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

```

```

RESULT 3
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

Db 29795 ATGGCTCTCGGCC 29807

## RESULT 4

US-09-141-908-1  
; Sequence 1, Application US/09141908  
; Patent No. 6503741  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: Combinatorial Polyketide Libraries' Produced Using a  
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold  
; FILE REFERENCE: 300622002100  
; CURRENT APPLICATION NUMBER: US/09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30  
; EARLIER APPLICATION NUMBER: PROV. 60/076,919  
; EARLIER FILING DATE: 1998-03-05  
; EARLIER APPLICATION NUMBER: PROV. 60/087,080  
; EARLIER FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-141-908-1

Query Match 31.7%; Score 158.6; DB 4; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 1.3e-20;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGGTCCCGACGCGGAGGAGACCGATCTCTGTAAGTGTGTCCGCGCCAGGCTTCG 60  
Db 29315 GCCTCTCCCGAGCCCGAGCGCGCGCTCTCTACCCCTCTGCTATCCCAACGCGCG 29374  
QY 61 ACGGTGCTCGGCGCACAGCGCGCGGAGGATCGGCCCGCGCAGCGGTTCACGAGGTC 120  
Db 29375 GCGTACTCGGCGCATCTCTCCCGACCGGTGCGCCCGCGCTGCTTACCGAGCTC 29434  
QY 121 GGTTCGACTCGTGGCGCGGTCAACTCCGCAACAGCTGCAACGCGGCGCAACGCGGCTG 180  
Db 29435 GGCTTCGACTCGTGAAGCGGTGCACTCCGCAACAGCTCTCCAGGTGGTGGCAAC 29494  
QY 181 CGGCTCGCGGAGCGGTGATCTTGGACTACCCACCGCGAGCGGTGGTGGTACTG 240  
Db 29495 AGGCTCCCGCCACACGCTTTCGACCCCGACCGCCCGGCTTCCGCGGCACTC 29554  
QY 241 CGGTGCAACTCTCTGCGGAGCGCGACGAGCGCTGACGCGGCGGGAAGACGACTCCGG 300  
Db 29555 CACGAGCGTACTCTCGCACCGCGCGGCGCGCGCGCGCGCGCGCGGTGGCG 29614  
QY 301 CGAGTCTCTCGCGCGGTGCGCGGTTCAGGAGGCGGCGGTGCTGGACACGCTG 360  
Db 29615 CGGCGCTTGGCGCAACTCGCCCTCGACCGGTGCGGAGCGCGGCGGTCTCGACACGCTC 29674  
QY 361 CTGGCTCTCGCGCGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 417  
Db 29675 CTGGCTCTCGCGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 29734  
QY 418 GCCCGCGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 477  
Db 29735 GACCTGTGCGGAGCGCGGAGCGGTGATCGACGACTTGGACATCTCGGTCTCGTCAA 29794  
QY 478 CGAGCGCTCTCGGCC 490  
Db 29795 ATGGCTCTCGGCC 29807

## RESULT 6

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328

## RESULT 5

US-09-657-440-19  
; Sequence 19, Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 31.7%; Score 158.6; DB 4; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 1.3e-20;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGGTCCCGACGCGGAGGAGACCGATCTCTGTAAGTGTGTCCGCGCCAGGCTTCG 60  
Db 29315 GCCTCTCCCGAGCCCGAGCGCGCGCTCTCTACCCCTCTGCTATCCCAACGCGCG 29374  
QY 61 ACGGTGCTCGGCGCACAGCGCGCGGAGGATCGGCCCGCGCAGCGGTTCACGAGGTC 120  
Db 29375 GCGTACTCGGCGCATCTCTCCCGACCGGTGCGCCCGCGCTGCTTACCGAGCTC 29434  
QY 121 GGTTCGACTCGTGGCGCGGTCAACTCCGCAACAGCTGCAACGCGGCGCAACGCGGCTG 180  
Db 29435 GGCTTCGACTCGTGAAGCGGTGCACTCCGCAACAGCTCTCCAGGTGGTGGCAAC 29494  
QY 181 CGGCTCGCGGAGCGGTGATCTTGGACTACCCACCGCGAGCGGTGGTGGTACTG 240  
Db 29495 AGGCTCCCGCCACACGCTTTCGACCCCGACCGCCCGGCTTCCGCGGCACTC 29554  
QY 241 CGGTGCAACTCTCTGCGGAGCGCGACGAGCGCTGACGCGGCGGGAAGACGACTCCGG 300  
Db 29555 CACGAGCGTACTCTCGCACCGCGCGGCGCGCGCGCGCGCGCGGTGGCG 29614  
QY 301 CGAGTCTCTCGCGCGGTGCGCGGTTCAGGAGGCGGCGGTGCTGGACACGCTG 360  
Db 29615 CGGCGCTTGGCGCAACTCGCCCTCGACCGGTGCGGAGCGCGGCGGTCTCGACACGCTC 29674  
QY 361 CTGGCTCTCGCGCGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 417  
Db 29675 CTGGCTCTCGCGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 29734  
QY 418 GCCCGCGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 477  
Db 29735 GACCTGTGCGGAGCGCGGAGCGGTGATCGACGACTTGGACATCTCGGTCTCGTCAA 29794  
QY 478 CGAGCGCTCTCGGCC 490  
Db 29795 ATGGCTCTCGGCC 29807



Db 4429 GGATTGACTCGTACGGCGGTGAGCTCCGACACCGCTCAAGAACGCCACCGGCTG 4488  
 QY 181 CGGTGCTCCGGGACGCTGATCTTGACTACACCCACCGGAGGCGCTGGTGGGCTACCTG 240  
 Db 4489 GCGCTCCCGGCACTCTGCTCTTCTGACTACCGACCCCGGACGCTGGCGGAGTTCTCTC 4548  
 QY 241 CGGTGCAACTCTCTCGGAGGCGGACGACGCGCTGAGCGGCGGGAAGACGACCTCCGG 300  
 Db 4549 CTCGCGGAGATCTCTGGGCGAGCAGCGCGGTGCGCGGAGAGCTTCCGCTGAGCGGCGG 4608  
 QY 301 CGAGTCTCTCGGCGCGGTGCGGCTTCCGCCG 329  
 Db 4609 GTCGACGAGGAGCCCGTCCGATGCTGG 4637

RESULT 9  
 US-08-804-227C-1  
 ; Sequence 1, Application US/08804227C  
 ; Patent No. 5876991  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DeHoff, Bradley S.  
 ; APPLICANT: Kuhstoss, Stuart A.  
 ; APPLICANT: Rostock, Paul R., Jr.  
 ; APPLICANT: Sutton, Kimberly L.  
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: THOMAS G. PLANT 1501  
 ; STREET: LILLY CORPORATE CENTER  
 ; CITY: INDIANAPOLIS  
 ; STATE: IN  
 ; COUNTRY: USA  
 ; ZIP: 46285

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII(DOS) Text only  
 CURRENT APPLICATION DATA:  
 FILING DATE: February 21, 1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Plant, Thomas, G.  
 REGISTRATION NUMBER: 35,784  
 REFERENCE/DOCKET NUMBER: X-8231  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 317-276-2459  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 43280 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 816..14234  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 14351..19945  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 20010..31199  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 31232..36067  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 36249..41774  
 US-08-804-227C-1

Query Match 27.1%; Score 135.8; DB 2; Length 43280;

Best Local Similarity 57.0%; Pred. No. 1.5e-16;  
 Matches 248; Conservative 0; Mismatches 187; Indels 0; Gaps 0;  
 QY 16 GAGCAGAACCGATCTCTGTAAGCTGGTCCGCGCCACGCTTCGACGCTGCTCGGCCAC 75  
 Db 35510 GAGCGCGAGGAGATCTGCTCGGCTGGTGGCGCCGGCATGTGGCCGCCGTACTCGGCCAC 35569  
 QY 76 AGCGCGCGCGAGGATCGGCCCGCGCCAGGGCGTTCCAGGAGGTCCGGCTTCGACTCGCTG 135  
 Db 35570 CCGGGGACCGCGGACATCGTCCCGACCGTGTCTTCAGGAGCTGGGGTTCAGTTGCTC 35629  
 QY 136 GCCGCGGTCAACCTCCGCAACAGCTGACGCGGCCACCGGGGCTGCGGCTGCGCGGACG 195  
 Db 35630 ACCGCGGTGAGCTGGCCCGGGCGCTGGCGCGGAGTGGCGACGGAAGCTGCCGCCGACG 35689  
 QY 196 CTGATCTTCGACTACCCACCCCGGAGGCGCTGGTGGCTAGCTCGCGCTCGAACTCCTG 255  
 Db 35690 CTGGTCTTCGACCATCCGACTCGCGCGCGCGCTGAAACACCTGGCGGAGTGTGTGACA 35749  
 QY 256 CGGAGGCGCGACGACGCGCTCGACGCGGCGGGAAGACGACCTCCGGCGAGTCTCTCGCGGC 315  
 Db 35750 CGCGCGCGCGGTCCCGCGCGCGCTCCCGCGGAGGAGGCGCGCGCGCGCGCGCG 35809  
 QY 316 GTGCGGTTCGCGCGGTTCAGGAGGCGCGGTGCTGGAACAGCTGCTCGGCTTCGCGGAC 375  
 Db 35810 GTGCGCTCGAACCGGTGAGGGAAGCGCGCTGCTGGAAGCACTGCTGGGCTTCGCGCG 35869  
 QY 376 ACCGCGACCGAACCGGCGCGGACGCGGAGACGACACCGAAGCGCGCGCGCGCGCGGAC 435  
 Db 35870 GACGAATCCGGGCGGACAAACCCCGCTAGCTGTGCGCGCTCGCGGCGACCCCGCGG 35929  
 QY 436 GCAGAACTGATCGAC 450  
 Db 35930 GAGGACCGGACGCG 35944

RESULT 10  
 US-08-804-227C-7  
 ; Sequence 7, Application US/08804227C  
 ; Patent No. 5876991  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DeHoff, Bradley S.  
 ; APPLICANT: Kuhstoss, Stuart A.  
 ; APPLICANT: Rostock, Paul R., Jr.  
 ; APPLICANT: Sutton, Kimberly L.  
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: THOMAS G. PLANT 1501  
 ; STREET: LILLY CORPORATE CENTER  
 ; CITY: INDIANAPOLIS  
 ; STATE: IN  
 ; COUNTRY: USA  
 ; ZIP: 46285  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII(DOS) Text only  
 CURRENT APPLICATION DATA:  
 FILING DATE: February 21, 1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Plant, Thomas, G.  
 REGISTRATION NUMBER: 35,784  
 REFERENCE/DOCKET NUMBER: X-8231  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 317-276-2459  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 44377 base pairs  
 TYPE: nucleic acid



QY 421 CGCGCGCGGACGACG 436  
 Db 36009 ACAGCGCGCGGACGCG 36024

RESULT 12  
 US-07-642-734C-1  
 ; Sequence 1, Application US/07642734C  
 ; Patent No. 5824513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Katz, L  
 ; APPLICANT: Donadio, S  
 ; APPLICANT: Mcalpine, J B  
 ; TITLE OF INVENTION: Recombinant DNA Method for Producing  
 ; TITLE OF INVENTION: Erythromycin Analogs  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Edward H. Gorman  
 ; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
 ; STREET: Park Rd  
 ; CITY: Abbott Park  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/642,734C  
 ; FILING DATE: 17-JAN-91  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dankers, Andreas M  
 ; REGISTRATION NUMBER: 32652  
 ; REFERENCE/DOCKET NUMBER: 4952.US.01  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 708-937-9396  
 ; TELEFAX: 708-938-2623  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 11219 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: DNA (genomic)  
 ; ANTI-SENSE: NO  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Saccharopolyspora erythraea  
 ; STRAIN: NRRL 2338  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 744..6659  
 ; OTHER INFORMATION: /function= "APPROXIMATE SPAN OF  
 ; OTHER INFORMATION: MODULE 1"  
 ; OTHER INFORMATION: /label= FUNCTION  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 744..11219  
 ; OTHER INFORMATION: /function= "gene= "eryA"  
 ; OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for  
 ; OTHER INFORMATION: 6-deoxyerythronolide B"  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 744..1868  
 ; OTHER INFORMATION: /function= "approximate span of  
 ; OTHER INFORMATION: acyltransferase domain 1 of module 1"  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 1998..2198

OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: acyl carrier domain 1 of module 1"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 2250..3626  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 3831..4811  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: acyltransferase 2 domain of module 1"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 5574..6125  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: beta-ketoreductase domain of module 1"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 6369..6626  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: acyl carrier domain 2 of module 1"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 6678..11219  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: module 2"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 6678..8066  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 8262..9305  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: acyltransferase domain of module 2"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 9906..10454  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: beta-ketoreductase domain of module 2"  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: 10707..10964  
 OTHER INFORMATION: /function= "approximate span of  
 OTHER INFORMATION: acyl carrier domain of module 2"  
 ;  
 US-07-642-734C-1

Query Match 26.1%; Score 130.8; DB 1; Length 11219;  
 Best Local Similarity 60.1%; Pred. No. 1.3e-15;  
 Matches 285; Conservative 0; Mismatches 162; Indels 27; Gaps 3;

QY 13 GCGAGCAGAACCGCATCTCTGCTGAAGCTGGTCGCGGCGCACGCTTCGACGCGTCTCGGC 72  
 Db 10725 GCGAGCGGACGCGGAGCTGGTGGCTCGCGACCGACCGGACCGCTGCTGGGC 10784

QY 73 CACAGGGCGCGAAGGATCGGCGCGCGCGCGGCTTCAGGAGGTCGCTTCGACTCG 132  
 Db 10785 CACAGCGCGGAGGCGGTGGCGCGCGCACCGCGCTTCAGGAGGTCGCTTCGACTCG 10844

QY 133 CTGGCGCGGTCAACCTTCGCAACAGCTTCGACGCGCGCACCGGCTTCGCGCTGCGCGCG 192  
 Db 10845 CTGGCGCGGTTCGCGGTGCGCAACCTTCGCAACCGGCGCGCGGCTTCGCGCTGCGCG 10904

QY 193 AGCTGATCTTCGACTACCCCGCGAGCGCGTGGTTCGCTACCTTCGCGCTCGAATC 252  
 Db 10905 AGCTGCTTCGACACCGCGCGCGCTTCGCGCGGTTCCTTCGCGCGGCTC 10964

QY 253 -----CTGCGGGAGGCGCGACGCGCTGCGCGGCGGAGAC 291  
 Db 10965 GGCACCGAGGTTCGCGGGGAGGCGCGCTTCGCGCGGCTGCGCGCTGGAAGGC 11024





10905	ACGCTGGTCTTCGACCAACCGCCTCCGGGTCGCGGGTTCTCTCGACGCCGAGTC	10964
	-----CTCGGGAGGCCGACGACGCGCTCGACGGCGGGAGAC	291
10965	GGCACCGAGGTCGGGGGAGGCGCCTCGGCCCTCGCGGGCTGACGCGCTGGAAGGC	11024
	-----GAGCGGGCGGTG---GAGCGGGCGGTG	348
11025	GGCCTCCCGAGGTGCCCAACCGACGGGGAAGAGCTGTCACGCGTCTGGAACGGATG	11084
	-----GACCGGCGCGGACACCGGACCGGACCGGACCGCGGAGACC	408
11085	CTCGCGCGCTAGGCCGGTGGCCAGGCCGCGGACGCGCTTCGGGACGCGCGCCAAACCG	11144
	-----GCCGACGACGCGAGAACTGATCGACGCACTGGAC	459
11145	TCCGGCGGACGACTGGGCGAGGCGGCGTGGACGAACTGCTCGAAGCACTGGC	11198

```

RESULT 14
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO.1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

```

Query Match	25.4%;	Score	127.4;	DB 3;	Length	50937;			
Best Local Similarity	62.3%;	Pred.	No. 4.8e-15;						
Matches	200;	Conservative	0;	Mismatches	121;	Indels	0;	Gaps	0;
Qy	5	TCCCCGACGCCGAGCAGAACCCTACTCTCAAGCTGGTTCGCGGCACACGTTCCGACGG	64						
Db	13102	TCACCCGGTTCGAGCAGGAACGTCTTCTGGTGAGCTCTGTAGGGAGCAGGCCCGCGTCTG	13161						
Qy	65	TGATCTCGGCCACACGCGCGCGCGAAGGCATCGGCCCGCGCCAGGCGCTTCAGGAGAGTCTGGCT	124						
Db	13162	TCCTCGSACAATTCCGGCGCGGAGGCGGTGCTTCGGACCGACGCTTCAAGGATCTCGAT	13221						
Qy	125	TCGACTCGCTTGGCGCGCGGTCAACCTCCGCAACAGCCTTGACCGCGGCACACGGGCTTGGCGGC	184						
Db	13222	TCGACTCGCTTGACCTCGGTTCGAACCTCGCAACCGGCTGAACACCCGCCACACGGCCCTCAGAC	13281						
Qy	185	TGCCCGGACGCTGANTCTTCGACTACCCACCCCGAGGCGCTGGTTCGGCTTACCTTCCGCG	244						
Db	13282	TGCCCGGTGACGGCGCGTCTTCGACTACGCGAGGCCCGCGCGCTTGGCGGCGCATCTGCGCT	13341						
Qy	245	TCGAACTCCTTCGCGGAGCGCCACGACGGCTGTGACGGGCGGGAAAGACACACTTCCGCGCAG	304						
Db	13342	CCAGGCTGATGACAGCAGCATGGTGTGACCAACGGTGCCTTGTCCCGCGTGTGAGAGCACGCGA	13401						
Qy	305	TCCTTCGCGCGCGTGCCTTCG	325						
Db	13402	TCGACGAGCCGATCGCGCATCG	13422						

```

RESULT 15
US-09-154-083-13/c
; Sequence 13, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 13
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-13

```

Query Match	25.2%	Score 126.2	DB 3	Length 561
Best local Similarity	67.0%	Pred. No. 9.4e-15		
Matches 179	Conservative 0	Mismatches 88	Indels 0	Gaps 0
Qy	16	GAGCAGAAACGGCATCTCTGTGTAAGCTGGTCCGGGGCCACAGCTTCCAGCGGTGCTCGGCCAC	75	
Db	398	GAGCGGCCGCACACACTCTGTGGAGCTGGTGGCGCACGAGCGCGATGCGGCTCGGCCAC	339	
Qy	76	AGCGGCGCCGAAAGGCATCGGCCCGCCGAGCGCTTCCAGGAGTTCGGCTTCGACTCGCGT	135	
Db	338	GGCTCCGAGCGGCGGTGGACGCGCAAGAGCGTTCGGGGAGATCGGCTTCGACTCGCGT	279	
Qy	136	GCOCGGTCAACTCTCCGCAACAGCGCTGCAGCGCGGCCACCGGGCTGCGGCTGCCCGCGACG	195	
Db	278	ACGCGCGTGAACCTGCGGAACCGGCTGGATACGGCAACGGGCGCTTAGGCTTCGCGCGACG	219	
Qy	196	CTGATCTTCGACTACCCACCCCGGAGCGCTGGTGGCTACTGCGGTCGGAATCTCGT	255	
Db	218	CTGGTGTTCGACTACCCGACCCCGCGGCGCTGGCGGGCGCGGCTGGGTGCGGAGCTTTC	159	
Qy	256	CGGGAGGCCGACGACGGCGCTGGACGG	282	
Db	158	GGCGGAGGCGGATGCGCTGGCGGCG	132	

Search completed: June 17, 2004, 03:22:09  
Job time : 68.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:31:09 ; Search time 293.667 Seconds  
(without alignments)  
7787.123 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcggtcccgagcgcgagca.....ccctcgggcagcagctga 501

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	158.6	31.7	4689	9 US-09-861-289-34	Sequence 34, Appl
2	158.6	31.7	4689	9 US-09-860-846-34	Sequence 34, Appl
3	158.6	31.7	4689	10 US-09-988-384B-34	Sequence 34, Appl
4	158.6	31.7	4689	10 US-09-836-821-34	Sequence 34, Appl
5	158.6	31.7	4689	15 US-10-271-889-34	Sequence 34, Appl
6	158.6	31.7	36778	9 US-09-861-289-5	Sequence 5, Appl
7	158.6	31.7	36778	9 US-09-860-846-5	Sequence 5, Appl
8	158.6	31.7	36778	15 US-10-271-889-5	Sequence 5, Appl
9	158.6	31.7	36778	15 US-10-271-889-48	Sequence 48, Appl
10	158.6	31.7	37948	10 US-09-988-384B-5	Sequence 5, Appl
11	158.6	31.7	38506	10 US-09-793-708-19	Sequence 19, Appl
12	158.6	31.7	38506	15 US-10-201-365-1	Sequence 1, Appl
13	158.6	31.7	38506	15 US-10-160-539-19	Sequence 19, Appl
14	155.8	31.1	4725	15 US-10-205-032-17	Sequence 17, Appl

15	155.8	31.1	60196	15	US-10-205-032-1	Sequence 1, Appl
16	153.6	30.7	5721	15	US-10-156-761-2880	Sequence 2880, Ap
17	153.6	30.7	125746	15	US-10-156-761-15102	Sequence 15102, A
18	153.6	30.7	9025608	15	US-10-156-761-1	Sequence 1, Appl
19	148.8	29.7	11238	15	US-10-205-032-15	Sequence 15, Appl
20	147.2	29.4	18438	15	US-10-156-761-2886	Sequence 2886, Ap
21	147.2	29.4	125746	15	US-10-156-761-15102	Sequence 15102, A
22	147.2	29.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
23	141	28.1	18435	15	US-10-156-761-412	Sequence 412, App
24	141	28.1	100000	15	US-10-156-761-15103	Sequence 15103, A
25	140.8	28.1	14055	15	US-10-156-761-2883	Sequence 2883, Ap
26	140.2	28.0	11220	9	US-09-861-289-32	Sequence 32, Appl
27	140.2	28.0	11220	9	US-09-860-846-32	Sequence 32, Appl
28	140.2	28.0	11220	10	US-09-988-384B-32	Sequence 32, Appl
29	140.2	28.0	11220	10	US-09-836-821-32	Sequence 32, Appl
30	140.2	28.0	11220	15	US-10-271-889-32	Sequence 32, Appl
31	139.6	27.9	18717	15	US-10-156-761-927	Sequence 927, App
32	139.6	27.9	30690	17	US-10-204-862A-1	Sequence 1, Appl
33	136.6	27.3	84428	17	US-10-229-148B-1	Sequence 1, Appl
34	135.8	27.1	11817	15	US-10-156-761-2884	Sequence 2884, Ap
35	135.8	27.1	23238	15	US-10-156-761-415	Sequence 415, App
36	134.4	26.8	64492	13	US-10-378-083-1	Sequence 1, Appl
37	134.2	26.8	3072	13	US-10-282-122A-25666	Sequence 25666, A
38	131.4	26.2	85692	17	US-10-461-194-1	Sequence 1, Appl
39	131.2	26.2	10692	15	US-10-156-761-414	Sequence 414, App
40	130.8	26.1	86941	17	US-10-461-194-2	Sequence 2, Appl
41	130	25.9	10839	15	US-10-156-761-2882	Sequence 2882, Ap
42	130	25.9	11910	15	US-10-156-761-2879	Sequence 2879, Ap
43	128.4	25.6	5505	15	US-10-156-761-413	Sequence 413, App
44	127.4	25.4	50937	10	US-09-808-880-1	Sequence 1, Appl
45	126.4	25.2	13494	15	US-10-156-761-2355	Sequence 2355, Ap

#### ALIGNMENTS

RESULT 1  
US-09-861-289-34  
; Sequence 34, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-34

Query Match	31.7%	Score 158.6;	DB 9;	Length 4689;
Best Local Similarity	59.0%	Pred. No. 1.1e-27;		
Matches 291;	Conservative	0;	Mismatches 199;	Indels 3; Gaps 1;
QY	1	GGGTCCTCCGAGCGCGAGCAGAACCGCATCTGCTGAGCTGCTCGCGGCGCACGCTTCG 60		
Db	4183	GGGTCCTCCGAGCGCGAGCGCGCGGCTCTCTACCTGCTCGTACCCACCGCGCG 4242		
QY	61	AGGTCCTGCGGCACAGCGCGCGCGCATCGCGCGCGCGCGCGCTTCCAGAGGTC 120		
Db	4243	GGCTACTGCGGCATCTCTCCCGCGCGCGCGCGCGCTTCCAGAGGTC 4302		
QY	121	GGCTTCGACTCGCTGCGCGCGCGCTCAACTCCGCAACAGCTCGACGCGCGCGCGCTG 180		

Db 4303 GGCTTGACTCGCTGACGGCGGTGAGTCCGCAACAGAGTCTCCACGGTGGTCCGCAAC 4362  
Qy 181 CGGTGTCGGCGAGCTGACTTTCGACTACCCACCCCGGAGGCGGTGTCGGTACTCTG 240  
Db 4363 AGGCTCCCGCCACACAGCGTCTTCGACACCCGACCGCCCGCGACTCGCGGCGCACTC 4422  
Qy 241 CGCTGCAACTCTCTGCGGGAGCGGACGACGGCTTGAACGGCGCGGGAAGACGACTCCGG 300  
Db 4423 CACGAGCGGTACTCTGCACCGCGCGAGCGCGCCCGGACGACTGGAGGGCGGTGGC 4482  
Qy 301 CGAGTCTTCGGCGCGTTCGCGCTTCGCGCGTTCGCGGTTCAAGAGGCGGCGTTCGGAACAGCTG 360  
Db 4483 CGGCGCTTCGCGGAACTCGCCCTCGACCGGTTCGCGGACGCGGGGTCCTCGACACCGTC 4542  
Qy 361 CTGGGCTTCGCGGACCGGACCGAACCGGCGGACGCGGCGTTCGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGGCTTCGCGGACCGGACCGAACCGGCGGACGCGGCGTTCGCGGCGGCGGCGGCGG 4602  
Qy 478 CGAGCCCTCGGGC 490  
Db 4663 ATGGCTCTCGGCC 4675

## RESULT 2

US-09-860-846-34  
; Sequence 34, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860.846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-34

Query Match 31.7%; Score 158.6; DB 9; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
Qy 1 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGTAAGTGTGTCCGCGGCGCACTTCG 60  
Db 4183 GCGCTCCCGAGCGCGGCGCGCGTCTCTCACTCCCTCGTCCGTACCCACGCGCG 4242  
Qy 61 ACGGTGCTCGGCCACAGCGGCGCGAGGCGATCGGCGCGCGCGCGCGCGCGCGCGGTC 120  
Db 4243 GCCGTACTCGGCCATTCTCCCGACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 4302  
Qy 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 180  
Db 4303 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCG 4362  
Qy 181 CGGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 240  
Db 4243 GCCGTACTCGGCCATTCTCCCGACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 4302  
Qy 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 180  
Db 4303 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCG 4362  
Qy 181 CGGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 240  
Db 4363 AGGCTCCCGCCACACAGCGTCTTCGACACCGGCGCGCGCGCGCGCGCGCGCGCG 4422  
Qy 241 CGGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 300  
Db 4423 CACGAGGCGTACCTCGCACCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4482

Qy 301 CGAGTCTTCGCGGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 360  
Db 4483 CGGCGCTTCGCGGAACTGCGCCCTCGACCGGTTCGCGGACGCGGCGGTTCCTCGACAC 4542  
Qy 361 CTGGGCTTCGCGGACCGGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGGCTTCGCGGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4602  
Qy 418 GCGCGGCGCGCGGACGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477  
Db 4603 GACCTGTCGCGGAGCGCGGCGGTTCGATCGACGCGGCGGCGGCGGCGGCGGCGG 4662  
Qy 478 CGAGCCCTCGGGC 490  
Db 4663 ATGGCTCTCGGCC 4675

## RESULT 3

US-09-988-384B-34  
; Sequence 34, Application US/0988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988.384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 34  
; LENGTH: 4689  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-34

Query Match 31.7%; Score 158.6; DB 10; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

Qy 1 GCGGTCCCGACGCGCGAGCAACCGCATCTCTGTAAGTGTGTCCGCGGCGCACTTCG 60  
Db 4183 GCGCTCCCGAGCGCGGCGCGCGTCTCTCACTCCCTCGTCCGTACCCACGCGCG 4242  
Qy 61 ACGGTGCTCGGCCACAGCGGCGCGAGGCGATCGGCGCGCGCGCGCGCGCGCGGTC 120  
Db 4243 GCCGTACTCGGCCATTCTCCCGACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 4302  
Qy 121 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCGCG 180  
Db 4303 GCGTTCGACTCGTGGCGCGGTCAACCTCGCAACAGCTTCGACGCGCGCGCGCGCG 4362  
Qy 181 CGGTTCGCGGAGCGTGTTCGACTACCCACCGCGGAGGCGGTGTCGGTACTCTG 240  
Db 4363 AGGCTCCCGCCACACAGCGTCTTCGACACCGCGCGCGCGCGCGCGCGCGCGCG 4422  
Qy 241 CGGTTCGAACTCTCTCGGGGAGCGGACGCGGCTTCGACGCGGCGGGAAGACGACTCC 300  
Db 4423 CACGAGGCGTACCTTCGACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4482  
Qy 301 CGAGTCTTCGCGGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTC 360  
Db 4483 CGGCGCTTCGCGGAACTGCGCCCTCGACCGGTTCGCGGACGCGGCGGTTCCTCGAC 4542  
Qy 361 CTGGGCTTCGCGGACCGGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 4543 CTGGGCTTCGCGGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4602

		GCCCCGGCCGACAGCAGCACTGATCGACACTTCGCGTCTCGTGCAA	477
Qy	418		
		GACCCCTGTGCGAGCGCGGGCTGATCGACACTCGCAGGCCCTGATCCGG	4662
Db	4603		
		CGAGCCCTCGGC	490
Qy	478		
		ATGGCTCTCGGC	4675
Db	4663		

```

RESULT 4
US-09-836-821-34
; Sequence 34, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

```

```

1 FILE OF INVENTION: DNA encoding metmyosin
2
3 TITLE REFERENCE: 600,438US1
4
5 CURRENT APPLICATION NUMBER: US/09/836,821
6
7 CURRENT FILING DATE: 2001-04-17
8
9 PRIORITY APPLICATION NUMBER: 09/105,537
10
11 PRIORITY FILING DATE: 1998-06-26
12
13 NUMBER OF SEQ ID NOS: 43
14
15 SOFTWARE: FastSeq for Windows Version 3.0
16
17 SEQ ID NO 34

```

; LENGTH: 4689  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-836-821-34

Query Match 31.7%; Score 158.6; DB 10; Length 4689;  
Best Local Similarity 59.0%; Pred. No. 1.1e-27;  
Matches 291; Conservative 0; Mismatches 199; Indels 3;

Qy	1	GGCGTCCCGACGCCGAGAGAAACCGCATCTCTGCTGAAGCTGTGTCCGCGCCACGCTTCG	60
Db	4183	GGCTCTCCCGACGCCGAGCCGCGCGCTCTCTACCCCTCGTTCGATACCCACGCGGGCG	4242
Qy	61	ACGGTGTCTCGGCGACACGGCGCCGGAAGGATCGGCCCGCGCCAGGCGTTCACGAGAGTC	120
Db	4243	GCGGTACTCGGCGCATCTCTCCCGACCGGGTGGCCCGCGCGTGTCTTCAACGAGCTC	4302
Qy	121	GGCTTCGACTCGCTGGCGCGGTCAACCTTCGCAAACGCTGCACGCGGCCACCGGGCTG	180
Db	4303	GGCTTCGACTCGCTGAACGCGCGTGCAGTTCGCAACACAGCTCTCCACGGTGGTGGGCAAC	4362
Qy	181	CGGCTGCCCGACGCTGATCTTCGATACCCCAACCCCGAGGCGCTGTGTCTGGCTACTGT	240
Db	4363	AGGCTCCCGCACCAACGGCTTCGACCAACGACGCGCCGCGCATCTCGCGCGCACCTC	4422
Qy	241	CGCGTCGAACCTCTTCGGGAGGCCGACGACGCGCTTCGACGCGCGGGGGAAGACGACCTCCGG	300
Db	4423	CACGAGGCGTACCTTCGACCGCGCGAGCGCGCCCGACGGAAGTGGGAGGGCGGGTCCGC	4482
Qy	301	CGAGTCTCTCGCGGCGCTGTCGGTTCGCGCGTTCAGGAGGCGGCGCTGTGTGACACGCTG	360
Db	4483	CGGGCCCTTGGCCGAACCTGCCCTTCGACCGGCTGGGGACGCGGGGTCTCTGACACCGTC	4542
Qy	361	CTCGGCTCTCGCGACACCGCGCACCGAAACGGGCGACGGACG---CCGAGACCCACCGAAGCG	417
Db	4543	CTGCGCTTCAACGGCATCGAGCCCGAGCCGAGTTCGGCGGTTTCGACGGCGGCGCCGCC	4602
Qy	418	GCCCCGGCCGCGACGACGCGAGAACTGATCGACGACATGGAACATCTCCGCTCTCGTGGAA	477
Db	4603	GACCTCTGTGCGGACCGGAGCGTTCGATCGACGACCTGGACGCGCGAGGCCCTGATCCGG	4662
Qy	478	CGAGCCCTCGGGC	490
Db	4663	ATGGCTCTCGGCC	4675

RESULT 5  
US 10-271-889-34  
; Sequence 34, Application US/10271889  
; Publication No. US20030194784A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.

```

1  APPLICANT: Zhao, L.
2  TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
3  FILE REFERENCE: 600.5820S1
4  CURRENT APPLICATION NUMBER: US/10/271.889
5  CURRENT FILING DATE: 2002-10-15
6  PRIOR APPLICATION NUMBER: US 09/861,289
7  PRIOR FILING DATE: 2001-05-18
8  PRIOR APPLICATION NUMBER: US 09/860,846
9  PRIOR FILING DATE: 2001-05-18
10 PRIOR APPLICATION NUMBER: US 09/836,821
11 PRIOR FILING DATE: 2001-04-17
12 PRIOR APPLICATION NUMBER: US 09/105,537
13 PRIOR FILING DATE: 1998-06-26
14 NUMBER OF SEQ ID NOS: 55
15 SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-34

```

Query Match	31.7%	Score 158.6;	DB 15;	Length 4689;
Best Local Similarity	59.0%;	Pred. No. 1.1e-27;		
Matches 291:	Conservative	0;	Mismatches 199;	Indels 3;

Qy	1	CGGGTCCCGAGCGCGGAGCAGAAACCGCATCTCTGCTGAAGCTGCTCCGCGGCGACGCTTCG	60
Db	4183	GGGCTCCCGAGCGCGGAGCGCGCGCGCTCTCTACCCCTCGTTCGTACCCACCGGGCG	4242
Qy	61	ACGGTGTCTGGCCACACAGCGCGCGGAGGCGATCGGCCCGCGCCAGGCGTTTCAGGAGGTC	120
Db	4243	GCGGTACTCTGGCCCATTTCTTCCCGGACCCGGGTGGCCCCCGCGCTGTCTTTACACGAGCTC	4302
Qy	121	GGCTTTCGACTCGCTGGCGCGGCTCAACTCCGCAACAGCCTGCAACGCGGCCACCGGGCTG	180
Db	4303	GGCTTTCGACTCGCTGACCGGCGGTGAGTTCGGCAACCGAGCTCTCCACGGTGGTCGGCAAC	4362
Qy	181	CGGCTGCCCGGACGACTGATCTTCGACTATCCCAACCCCGAGGCGCTGGTTCGGCTACTG	240
Db	4363	AGGCTCCCGGCGCACACGGTCTTCGACCAACCGACGCGCGCGCACTCGCGCGCACCTC	4422
Qy	241	CGGCTGGAATTCCTGGCGGAGCGCGACGACGCGCTTGAACGGGCGGGAAGACGACTCCGG	300
Db	4423	CACGAGGCGTACTTCGCAACCGCGCGAGCCGCGCCCGGACCGGACTGGGAGGGCGGGTCGC	4482
Qy	301	CGAGTCTCTCGGGCGGTGCGCTTCGCGCGGTTCAGAGAGCGGCGCTGCTGGACACGCTG	360
Db	4483	CGGGCCCTGGCGGAACCTGCCCTCGAACCGGTGCGGACGCGGGGGTCTTCGACACCGTC	4542
Qy	361	CTCGGCTCTGGCGACACCGGACCGCAACCGGCGACGGAACG---CCGAGACCAACCGAAGCG	417
Db	4543	CTGGCCCTACCCGGCATTCGAGCCCGAGCGGGTTCCGGCGGTTCGGACGGGGCGCGCC	4602
Qy	418	GCCCGGGCGCGGACGACGCGAGAACTGATCGAACCACTGGACATCTCCGGTCTCGTGCAA	477
Db	4603	GACCTGGTGGGAGCCGAGGGCGTTCGATCGAAGCACTGGACCGCGAGGCCCTGATCCGG	4662
Qy	478	CGAGCCCTCGGGC	490
Db	4663	ATGGCTCTCGGCC	4675

RESULT 6  
US-09-861-289-5

```
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match      31.7%; Score 158.6; DB 9; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCTCCCGACGCGGAGCAGAACCGCATCTCTGCTGAAGCTGTCTCCGCGGCGACGCTTCG 60
Db 31173 GCGGTCTCCCGAGCCCGAGCGCGCGCGCTCTCTACCTCTCTGTCCTGATCCACCGCGCG 31232

QY 61 ACGGTCTCGGCCACAGCGCGCGCGAGCATCGGCCCGCGCGCGAGGCTTCCAGGAGTC 120
Db 31233 GCGGTACTTCGGCCATTCTCTCCCGAGCCGGGTGGCCCCCGCGCTTCTTACCGAGCTC 31292

QY 121 GGCTTCGACTCGCTTGGCGCGGTCAACCTCCGCAACAGACCTGACGCGGCCACCGGGCTG 180
Db 31293 GGCTTCGACTCGCTGAGCGCGGTGCGAGCTCGCAACAGCTCTCCACGGTGGTGGCAAC 31352

QY 181 CGGTGTCGCGGACGCTGTATCTTCGACTACCCACCCCGAGGCGCTGTGCTGGCTACCTG 240
Db 31353 AGGCTCCCGCCACACAGCTTTCGACACCCGCGCGCGCTTCGCGCGCGCGACCTC 31412

QY 241 CGGTTCGAACTCTGCGGAGCGCGAGCGGCTTGAACGGGCGGGAAGACGCTCGG 300
Db 31413 CACGAGCGGTACTCTGCAACCGCGCGAGCCCGCGCGCGCTCTCCACGGTGGTGGCAAC 31352

QY 181 CGGTGTCGCGGACGCTGTATCTTCGACTACCCACCCCGAGGCGCTGTGCTGGCTACCTG 240
Db 31353 AGGCTCCCGCCACACAGCTTTCGACACCCGCGCGCGCTTCGCGCGCGCGACCTC 31412

QY 241 CGGTTCGAACTCTGCGGAGCGCGAGCGGCTTGAACGGGCGGGAAGACGCTCGG 300
Db 31413 CACGAGCGGTACTCTGCAACCGCGCGAGCCCGCGCGCGCTCTCCACGGTGGTGGCAAC 31352

QY 301 CGAGTCTCTCGCGCGCTGTGCGCTTTCGCGCGCGCTTCAAGAGCGCGGCGCTGTGACACGCTG 360
Db 31473 CGGGCCCTCGCGGAACTCTGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCT 31532

QY 361 CTGGGCTCTCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 417
Db 31533 CTGGGCTCTCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 31592

QY 418 GCCCGCGCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 477
Db 31593 GACCTCTGTCGCGGACCGGAGCGCTGATCGAGCACTGACGACATCTCGGCTCTGTCGAA 31652

QY 478 CGAGCCCTCGGCG 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 7
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

```
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match      31.7%; Score 158.6; DB 9; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCTCCCGACGCGGAGCAGAACCGCATCTCTGCTGAAGCTGTCTCCGCGGCGACGCTTCG 60
Db 31173 GCGGTCTCCCGAGCCCGAGCGCGCGCGCTCTCTACCTCTCTGTCCTGATCCACCGCGCG 31232

QY 61 ACGGTCTCGGCCACAGCGCGCGCGAGCATCGGCCCGCGCGCGAGGCTTCCAGGAGTC 120
Db 31233 GCGGTACTTCGGCCATTCTCTCCCGAGCCGGGTGGCCCCCGCGCTTCTTACCGAGCTC 31292

QY 121 GGCTTCGACTCGCTTGGCGCGGTCAACCTCCGCAACAGACCTGACGCGGCCACCGGGCTG 180
Db 31293 GGCTTCGACTCGCTGAGCGCGGTGCGAGCTCGCAACAGCTCTCCACGGTGGTGGCAAC 31352

QY 181 CGGTGTCGCGGACGCTGTATCTTCGACTACCCACCCCGAGGCGCTGTGCTGGCTACCTG 240
Db 31353 AGGCTCCCGCCACACAGCTTTCGACACCCGCGCGCGCTTCGCGCGCGCGACCTC 31412

QY 241 CGGTTCGAACTCTGCGGAGCGCGAGCGGCTTGAACGGGCGGGAAGACGCTCGG 300
Db 31413 CACGAGCGGTACTCTGCAACCGCGCGAGCCCGCGCGCGCTCTCCACGGTGGTGGCAAC 31352

QY 301 CGAGTCTCTCGCGCGCTGTGCGCTTTCGCGCGCGCTTCAAGAGCGCGGCGCTGTGACACGCTG 360
Db 31473 CGGGCCCTCGCGGAACTCTGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCT 31532

QY 361 CTGGGCTCTCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 417
Db 31533 CTGGGCTCTCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 31592

QY 418 GCCCGCGCGCGGACGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 477
Db 31593 GACCTCTGTCGCGGACCGGAGCGCTGATCGAGCACTGACGACATCTCGGCTCTGTCGAA 31652

QY 478 CGAGCCCTCGGCG 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 8
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
```

```
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match      31.7%; Score 158.6; DB 10; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGTCCCGGAGCGGAGCAGACCGCATCTGCTGAAGCTGTTCCGCGGCCACGCTTCG 60
Db 31173 GCGTCCCGGAGCGGAGCGGCGCGGCTCTCAACCTCGTCCGTTACCCACGCGCG 31232

QY 61 ACGGTGCTCGGCGCACAGCGGCGCGGAGCGCATCGCCCGCGCCAGCGCTTCCAGAGGTC 120
Db 31233 GCGGTACTCGGCGCATTCCTCCCGCGACCGGTTGGCCCGCGGCTTCAACGAGCTC 31292

QY 121 GCGTTGCACTCGTGGCGCGGTCAACCTCGCAACAGCCTGCGACCGCGCCACCGGGCTG 180
Db 31293 GCGTTGCACTCGTGGCGCGGTGCGAGCTTCGCAACAGCTCTCCACGGTGGTCCGCAAC 31352

QY 181 GCGTTCGCGGAGCGCTGATCTTCACTTACCCCGCGGAGCGCTGGTTCGGTACCTG 240
Db 31353 AGGCTCCCGCGCCACCGAGTCTTTCGACCGCGCGGCGGCTTCGCGCGCACCTC 31412

QY 241 CGGTCGCAACTCTCGGCGGAGCGCGACGAGCGCTTGAAGCGGCGGAGAGACGACCTCGG 300
Db 31413 CACGAGGCGTACTTCGCAACCGCGCGGAGCGCGCGCGGAGCGGCGGCTGCGC 31472

QY 301 CGAGTCTCTCGGCGCGTTCGCGGCTTCAAGAGGAGCGGCGGCTGTGACACGCTG 360
Db 31473 CGGCGCTTGGCGCACTGCGGCTTCAAGAGGAGCGGCGGCTGTGACACGCTG 31532

QY 361 CTCGCGCTTCGCGGAGCGGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 31533 CTGCGCTTCACCGGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31592

QY 418 GCGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 477
Db 31593 GACCTCTGTGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 31652

QY 478 CGAGCGCTTCGGGC 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 10
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.5360S1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match      31.7%; Score 158.6; DB 10; Length 37948;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGGAGCGGAGCAGAACCGCATCTGCTGAAGCTGTTCCGCGGCCACGCTTCG 60
Db 32343 GCGTCCCGGAGCGGAGCGGCGGCGGCTCTCAACCTCGTCCGTTACCCACGCGCG 32402

; ORGANISM: Streptomyces venezuelae
US-10-271-889-48

Query Match      31.7%; Score 158.6; DB 10; Length 36778;
Best Local Similarity 59.0%; Pred. No. 6.8e-28;
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGTCCCGGAGCGGAGCAGACCGCATCTGCTGAAGCTGTTCCGCGGCCACGCTTCG 60
Db 31173 GCGTCCCGGAGCGGAGCGGCGCGGCTCTCAACCTCGTCCGTTACCCACGCGCG 31232

QY 61 ACGGTGCTCGGCGCACAGCGGCGCGGAGCGCATCGCCCGCGCCAGCGCTTCCAGAGGTC 120
Db 31233 GCGGTACTCGGCGCATTCCTCCCGCGACCGGTTGGCCCGCGGCTTCAACGAGCTC 31292

QY 121 GCGTTGCACTCGTGGCGCGGTCAACCTCGCAACAGCCTGCGACCGCGCCACCGGGCTG 180
Db 31293 GCGTTGCACTCGTGGCGCGGTGCGAGCTTCGCAACAGCTCTCCACGGTGGTCCGCAAC 31352

QY 181 GCGTTCGCGGAGCGCTGATCTTCACTTACCCCGCGGAGCGCTGGTTCGGTACCTG 240
Db 31353 AGGCTCCCGCGCCACCGAGTCTTTCGACCGCGCGGCGGCTTCGCGCGCACCTC 31412

QY 241 CGGTCGCAACTCTCGGCGGAGCGCGACGAGCGCTTGAAGCGGCGGAGAGACGACCTCGG 300
Db 31413 CACGAGGCGTACTTCGCAACCGCGCGGAGCGCGCGCGGAGCGGCGGCTGCGC 31472

QY 301 CGAGTCTCTCGGCGCGTTCGCGGCTTCAAGAGGAGCGGCGGCTGTGACACGCTG 360
Db 31473 CGGCGCTTGGCGCACTGCGGCTTCAAGAGGAGCGGCGGCTGTGACACGCTG 31532

QY 361 CTCGCGCTTCGCGGAGCGGCGGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 417
Db 31533 CTGCGCTTCACCGGCGGAGCGGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAG 31592

QY 418 GCGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 477
Db 31593 GACCTCTGTGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 31652

QY 478 CGAGCGCTTCGGGC 490
Db 31653 ATGGCTCTCGGCC 31665

RESULT 9
US-10-271-889-48
; Sequence 48, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-48
```

QY 61 ACGTGCTCGGCACACGCGCGCGAGGATCGGCGCGCGCGAGGCTTCCAGAGGTC 120  
Db 32403 GCGGTACTCGGCATTTCTCTCCCGACCGGGTGGCCCGCGCGGTCTTACCGAGCTC 32462  
QY 121 GGCTTCGACTCGTGGCGCGGTCAACTCGCAACAGCTGACGCGGCAACCGGGCTG 180  
Db 32463 GGCTTCGACTCGTGGCGCGGTGAGCTCGCAACAGCTCTCCAGGTGTCGGAAC 32522  
QY 181 CGGTGCGCGGACGCTGACTTCGACTACCCACCGCGGAGGCGCTGCTGCTACTCG 240  
Db 32523 AGGTCTCCCGCCACACAGGCTTTCGACACCCGACCGCGCGGCTGCTGCTGCTGCT 32582  
QY 241 CGCTCGAACTCTCTCGGCGGAGCGACGACGCGCTGACGCGCGGAGAGCACTCCGG 300  
Db 32583 CACGAGGCTTACTTCGACCGCGCGGAGCGCGCGCGCGGAGGCGGCTGCGC 32642  
QY 301 CGATCTCTCGGCGCGGTGCGCTTCCCGGTTTCAAGAGGCGGCGGTGCTGGAACGCTG 360  
Db 32643 CGGCGCTTGGCGAACTGCGCTCGACCGGCTGCGGAGCGCGGCGGTCTCGACACGCTC 32702  
QY 361 CTCGGCTCGCGACACCGGACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 32703 CTGCGCTTACCGGATCGAGCGCGGCTTCCGCGGCTTCCGAGCGGCGGCGGCGGCG 32762  
QY 418 GCCCGCGCGCGGACGAGCACTGATCGACGCACTTGGACATCTCCGGTCTCGTGCAA 477  
Db 32763 GACCTGTGCGGAGCGGAGCGGTGATCGACGACCTGGAGCGCGGCGGCGGCGGCGG 32822  
QY 478 CGAGCCCTCGGCG 490  
Db 32823 ATGGCTCTCGGCC 32835

## RESULT 11

US-09-793-708-19

; Sequence 19, Application US/09793708  
; Publication No. US20030104597A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; CURRENT APPLICATION NUMBER: US/09/793,708  
; CURRENT FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: US/09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: US/09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: US/09/073,538  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US/08/846,247  
; PRIOR FILING DATE: 1997-04-30  
; PRIOR APPLICATION NUMBER: US/60/134,990  
; PRIOR FILING DATE: 1999-05-20  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-793-708-19

Query Match 31.7%; Score 158.6; DB 10; Length 38506;

Best Local Similarity 59.0%; Pred. No. 6.7e-28;

Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGACGCGGAGCAACCGCATCTCTGTAAGCTGCTCGGCGGCGGCGGCTCG 60

Db 29315 GCGTCTCCCGAGCCCGAGCGCGCGCGCTCTCTCACTCTGCTCCGACGCGGCG 29374  
QY 61 ACGGTGCTCGGCACACGCGCGCGAGGATCGGCGCGCGCGAGGCTTCCAGAGGTC 120  
Db 29375 GCGGTACTCGGCATTTCTCTCCCGACCGGGTGGCCCGCGGTGCTTACCGAGCTC 29434  
QY 121 GGCTTCGACTCGTGGCGCGGTCAACTCGCAACAGCTGACGCGGCAACCGGGCTG 180  
Db 29435 GGCTTCGACTCGTGGCGCGGTGAGCTCGCAACAGCTCTCCAGGTGTCGGAAC 29494  
QY 181 CGGTGCGCGGACGCTGACTTCGACTACCCACCGCGGAGGCGCTGCTGCTACTCG 240  
Db 29495 AGGTCTCCCGCCACACAGGCTTTCGACACCCGACCGCGGCTGCTGCTGCTGCT 29554  
QY 241 CGCTCGAACTCTCTCGGCGGAGCGACGACGCGCTGACGCGGCGGAGAGCACTCCGG 300  
Db 29555 CACGAGGCTTACTTCGACCGCGCGGAGCGCGCGCGGAGGCGGCTGCGC 29614  
QY 301 CGATCTCTCGGCGCGGTGCGCTTCCCGGTTTCAAGAGGCGGCGGTGCTGGAACGCTG 360  
Db 29615 CGGCGCTTGGCGAACTGCGCTCGACCGGCTGCGGAGCGCGGCGGTCTCGACACGCTC 29674  
QY 361 CTCGGCTCGCGACACCGGACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 29675 CTGCGCTTACCGGATCGAGCGCGGCTTCCGCGGCTTCCGAGCGGCGGCGGCGGCGG 29734  
QY 418 GCCCGCGCGCGGACGAGCACTGATCGACGCACTTGGACATCTCCGGTCTCGTGCAA 477  
Db 29735 GACCTGTGCGGAGCGGAGCGGTGATCGACGACCTGGAGCGCGGCGGCGGCGGCGG 29794  
QY 478 CGAGCCCTCGGCG 490  
Db 29795 ATGGCTCTCGGCC 29807

## RESULT 12

US-10-201-365-1

; Sequence 1, Application US/10201365  
; Publication No. US20030148469A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR  
; TITLE OF INVENTION: PKS GENE CLUSTER AS SCAFFOLD  
; FILE REFERENCE: 300622002103  
; CURRENT APPLICATION NUMBER: US/10/201,365  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: US/09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: US/09/073,538  
; PRIOR FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-10-201-365-1

Query Match 31.7%; Score 158.6; DB 15; Length 38506;

Best Local Similarity 59.0%; Pred. No. 6.7e-28;

Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 1 GCGGTCCCGACGCGGAGCAACCGCATCTCTGTAAGCTGCTCGGCGGCGGCGGCTCG 60

Db 29315 GCGTCTCCCGAGCCCGAGCGCGCGCGGTCTCTCTCACTCTGCTCCGTAACCGCGGCG 29374

QY 61 ACGGTGCTCGGCACACGCGGCGGAGGATCGGCGCGCGGCGGCGGTCTCCAGAGGTC 120

Db 29375 GCGGTACTCGGCATTTCTCTCCCGAGCGGCTGCGGCGGCGGCGGCTTACCGAGGTC 29434

QY 121 GCTTCGACTCGCTGGCCGCGGTCAACCTCCGAAACAGCCTGCACGCGCCACCGGGCTG 180  
DB 29435 GCTTCGACTCGCTGACGCGCGGTGAGCTCCGCAACAGCTCTCCACGGTGTGCGCAAC 29494  
QY 181 CGGCTGCGCGGACGCTGATCTTGAATACCCACCCCGGAGGCGCTGGTGGCTACCTG 240  
DB 29495 AGGCTCCCGCGCACACAGGTCTTGAACACCGGACCGCGCGGCGACTCGCGCGCACCTC 29554  
QY 241 CGGCTGCAACTCTCGGGAGCGCGACGACGCGCTGAGAGCGGGCGGGAAGACACCTCCG 300  
DB 29555 CACGAGGCGTACTCTCGACCGCGCGGCGCGGCGCGGCGGCGGCGGCTGCGC 29614  
QY 301 CGAGTCTCTCGGCGCGGTGCGCTTCCCGGTTCAAGGAGGCGGCGGTCTGACACGCTG 360  
DB 29615 CGGCGCTTGGCGCAACTGCCCTCGACCGGCTGCGGAGCGGGGTCTCTGACACCGTC 29674  
QY 361 CTCGGCTTGGCGGACCGGACCGGACCGGCGGCGGAGCG---CGAGACACCGGAAGCG 417  
DB 29675 CTGCGCTTACCGGATCGAGCCGCGGCTTCCGCGGTTCGAGCGCGCGCGCC 29734  
QY 418 GCCCGGCGCGGACGACGAGAACTGATCGACGCACTGAGACATCTCGGTCTGTGCAA 477  
DB 29735 GACCTGTGCGGAGCGGAGCGGTGATCGACGACCTGAGCGGCGCTGATCCGG 29794  
QY 478 CGAGCCCTCGGCG 490  
DB 29795 ATGCTCTCGGCC 29807

## RESULT 13

US-10-160-539-19  
; Sequence 19, Application US/10160539  
; Publication No. US20030162262A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/10/160,539  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-10-160-539-19

Query Match 31.7%; Score 158.6; DB 15; Length 38506;  
Best Local Similarity 59.0%; Pred. No. 6.7e-28;  
Matches 291; Conservative 0; Mismatches 199; Indels 3; Gaps 1;  
QY 1 GCGGTCCCGACCGCGGAGCAGAACCGCATCTCTGTAAGCTGTCTCGCGGCGCACCGCTTCG 60  
DB 29315 GCGTCCCGGAGCGCGGCGCGGCGGTCTCTCACCTCTGTCGTAACCGCGCGG 29374  
QY 61 AGGTGCTCGGCGACAGCGCGCGGAGGATCGGCGCGGCGGAGCGGTTCAGAGAGTTC 120  
DB 29375 GCGTACTCGGCGATCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 29434  
QY 121 GCTTCGACTCGCTGCGCGCGGTCAACTCCGCAACAGCTTGCACGCGGCGCACCGGGCTG 180  
DB 29435 GCTTCGACTCGCTGACGCGCGGTGAGCTCCGCAACAGCTCTCCACGCTGTGTCGCGCAAC 29494

QY 181 CGGCTGCGCGGAGCAGCTGATCTTGAATACCCACCCCGGAGGCGGTGGTGGCTACCTG 240  
DB 29495 AGGCTCCCGCGCACACAGGTCTTTCGACACCGGACCGCGCGGCGACTCGCGCGCACCTC 29554  
QY 241 CGGCTGCAACTCTCGGGAGCGCGACGACGCGCTTGAAGCGGGGAAGACACCTCCG 300  
DB 29555 CACGAGGCGTACTCTCGACCGCGCGGCGCGGCGCGGCGGCGGCGGCTGCGC 29614  
QY 301 CGAGTCTCTCGGCGCGGTGCGCTTCCCGGTTCAAGGAGGCGGCGGTCTGACACGCTG 360  
DB 29615 CGGCGCTTGGCGCAACTGCCCTCGACCGGCTGCGGAGCGGGGTCTCTGACACCGTC 29674  
QY 361 CTCGGCTTGGCGGACCGGACCGGACCGGCGGCGGAGCG---CGAGACACCGGAAGCG 417  
DB 29675 CTGCGCTTACCGGATCGAGCCGCGGCTTCCGCGGTTCGAGCGCGCGCGCC 29734  
QY 418 GCCCGGCGCGGACGACGAGAACTGATCGACGCACTGAGACATCTCGGTCTGTGCAA 477  
DB 29735 GACCTGTGCGGAGCGGAGCGGTGATCGACGACCTGAGCGGCGCTGATCCGG 29794  
QY 478 CGAGCCCTCGGCG 490  
DB 29795 ATGCTCTCGGCC 29807

## RESULT 14

US-10-205-032-17  
; Sequence 17, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-2US  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 17  
; LENGTH: 4725  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-17

Query Match 31.1%; Score 155.8; DB 15; Length 4725;  
Best Local Similarity 62.2%; Pred. No. 4.7e-27;  
Matches 263; Conservative 0; Mismatches 157; Indels 3; Gaps 1;  
QY 12 CGCGGAGCAGAACCGCATCTCTGTAAGCTGTCTCGCGGCGCACCGCTTCGACGCTGCTCGG 71  
DB 4254 CGACGAAACCGCGCGGCGCTGCTCGACCTGTGTACCGCGGAGGTGCGCGGCTCTCTCG 4313  
QY 72 CCACGCGCGCGGAGGAGCATCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 131  
DB 4314 CCACCGCGGCGCGGAGCAGCTCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4373  
QY 132 GTGCGCGCGGCGCAACCTCCGCAACAGCTTGCAGCGGCGCACCGGCGGCGGCGGCGGCGG 191  
DB 4374 GTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4433  
QY 192 GACGTGATCTTCGACTACCCACCCCGGAGCGCTGTGCGGCTACCTCGGCGTGAAT 251  
DB 4434 CACCTCTGCTTCGACACACCCACCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4493  
QY 252 CCGCGGAGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 311  
DB 4494 CGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4550  
QY 312 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 371  
DB 4551 CGACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4610





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 23:18:23 ; Search time 1960 Seconds  
(without alignments)  
7633.139 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcgggtccgcagccgagca.....ccctcgggcagagctga 501

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gsa\_hum:\*  
18: em\_gsa\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	75	15.0	925	29	CNS0091P	AL053013 Drosophil
C 2	74	14.8	924	13	EX442207	EX442207 BX442207
C 3	72.8	14.5	924	13	EX442207	EX442207 BX442207
C 4	71.4	14.3	982	13	EX415111	EX415111 BX415111

5	71.2	14.2	925	29	CNS0091P	AL053013 Drosophil
c	71	14.2	935	29	CNS006XK	AL066051 Drosophil
c	70.8	14.1	1104	13	BQ722521	BQ722521 AGENCOURT
8	70	14.0	1232	29	AG072425	AG072425 Pan trogl
c	68.8	13.7	803	29	AG042898	AG042898 Pan trogl
10	68.8	13.7	932	29	CNS00720	AL066742 Drosophil
c	68.6	13.7	932	29	CNS0072Q	AL066742 Drosophil
12	67.6	13.5	935	29	CNS006XK	AL066051 Drosophil
13	66.8	13.3	622	14	CB654160	CB654160 OSJNEC050
14	66.4	13.3	1516	12	BG09984	BG09984 mgct002xd
15	66.2	13.2	562	13	BQ606671	BQ606671 BRY 2539
16	66	13.2	621	14	CF326124	CF326124 JMT1--05-
17	66	13.2	633	14	CF325834	CF325834 JMT1--04-
18	65.8	13.1	982	13	EX415111	EX415111 BX415111
19	65.8	13.1	1134	12	BM915656	BM915656 AGENCOURT
c	63.8	12.7	965	28	AQ893417	AQ893417 HS 4832 A
21	63.2	12.6	438	13	BQ805019	BQ805019 WHE3561_H
22	63	12.6	710	9	AU163439	AU163439 AU163439_H
c	63	12.6	1144	13	EX415926	EX415926 BX415926
24	62.6	12.5	638	28	BZ895164	BZ895164 HG8 0108
c	62.2	12.4	1094	29	AG036373	AG036373 Pan trogl
26	61.8	12.3	449	10	BF728957	BF728957 1000069G1
27	61.8	12.3	790	14	CB671171	CB671171 OSJNE804J
28	61.8	12.3	829	14	CB642787	CB642787 OSJNE803C
30	61.6	12.3	997	29	AG057392	AG057392 Pan trogl
c	61.4	12.3	805	29	CG287915	CG287915 OGICX89TV
31	61.4	12.3	840	29	CG327544	CG327544 OGXBDA47TV
32	61.2	12.2	615	14	CB627718	CB627718 OSIIEB02N
33	61.2	12.2	688	29	CG263720	CG263720 OGYCO36TH
34	61.2	12.2	771	14	CB670011	CB670011 OSJNE802N
35	61.2	12.2	1101	29	CNS0175X	AL08460 Drosophil
36	61	12.2	653	13	BU673539	BU673539 NL 16 80
37	60.8	12.1	718	14	CD868844	CD868844 AZ02.109P
c	60.8	12.1	832	29	CC719244	CC719244 OGWC751TH
c	60.8	12.1	841	29	CG291746	CG291746 OG2B094TH
40	60.8	12.1	918	29	CC719255	CC719255 OGWC751TV
41	60.8	12.1	959	29	CG341580	CG341580 OGTBU16TV
42	60.8	12.1	1112	14	CA477644	CA477644 AGENCOURT
c	60.8	12.1	1122	12	BM914385	BM914385 AGENCOURT
44	60.6	12.1	738	14	CF868028	CF868028 tric013xh
45	60.6	12.1	810	14	CB898085	CB898085 tric013xh

## ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION  
AL053013  
VERSION  
AL053013.1 GI:4934461  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 925)  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
JOURNAL  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of







**AUTHORS**  
**TITLE** NTH-MGC <http://mgc.nci.nih.gov/>.  
**JOURNAL** National Institutes of Health, Mammalian Gene Collection (MGC)  
**COMMENT** Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM13583 row: 9 column: 03  
 High quality sequence stop: 297.

**FEATURES**  
 source  
 1..1104  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6188186"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski syphathetic trunk"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1: NciI; Site\_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCACGCTCCG-3' and 5'-GACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

## ORIGIN

Query Match 14.1%; Score 70.8; DB 13; Length 1104;  
 Best Local Similarity 47.0%; Pred. No. 3.2;  
 Matches 215; Conservative 0; Mismatches 241; Indels 1; Gaps 1;

Qy 39 GCTGGTCCGGGCCACGCTTCGACGGTGTCTGGCCACAGCGCGGCCGGAAGGATCGGCC 98  
 |||  
 Db 1069 GCGGACGCGCCGCGCCG 1010

Qy 99 GCGGACGCGGTTCAGAGGTTCGGCTTCGACTCGCTGGCGCGCGGTCAACTTCGCAACAG 158  
 |||  
 Db 1009 GCCCGCGGGTCTCG 950

Qy 159 CTTGACGCGCGCCACCGGGTTCGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 218  
 |||  
 Db 949 GGCNCGMCGCGCCCG 890

Qy 219 GGAGGCGCTGTCTGGCTTACTTCGCGGTTCGAACTCTCGGGAGGCGCGACGACGCGCTTGA 278  
 |||  
 Db 889 CGCGGCG 830

Qy 279 CGGCGCGGAGAGACGACTTCGCGGAGTCTCTCGCGCGCGGTGCGTTCGCGCGGTTCAGGA 338  
 |||  
 Db 829 CGCCG 770

Qy 339 GCGGCGGTGTGTGACACGCTCTCGGCTTCGCGGACACCGGACCGGACCGGACCGGACGGA 398  
 |||  
 Db 769 CCCGCGGCG 710

Qy 399 CGCGGAGACACCGAAGCG 458  
 |||  
 Db 709 CCCNCCCG 495

Qy 459 CATCTCGGCTCTGTGCAACGAGCCCTCGGGCAGACG 495  
 |||  
 Db 649 CGGC-CG 614

## RESULT 8

AG072425  
 LOCUS Pan troglodytes DNA, clone: PTB-063017.F, genomic survey sequence.  
 AG072425  
 DEFINITION AG072425  
 ACCESSION AG072425  
 VERSION 1 GI:16624227  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of Library PTB  
 Unpublished  
 2 (bases 1 to 1232)  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
 Direct Submission  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimpanzee@gsc.riken.go.jp](mailto:chimpanzee@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

## PRIMERS

Sequencing: -21M13

## LIBRARY

Vector : pXS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1..1232

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

/clone="PTB-063017.F"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 14.0%; Score 70; DB 29; Length 1232;  
 Best Local Similarity 47.1%; Pred. No. 4.1;  
 Matches 210; Conservative 0; Mismatches 235; Indels 1; Gaps 1;

Qy 42 GGTCCGCGGCCACGCTTCGACGGTGTCTGGCCACAGCGGCGGCCGCAAGGATCGGCCGCG 101  
 |||  
 Db 535 GGGCG 594

Qy 102 CCAGGCGTTCAGAGGTTCGGCTTCGACTCGCTGGCGCGGTCAACTTCGCAACAGCGCT 161  
 |||  
 Db 595 CGGCGGCG 654

Qy 162 GCACGCGCGCCACCGGGCTGCGGCTGCGCCGCGACGCTGATCTTCGACTACCCACCCCGGA 221  
 |||  
 Db 655 GCCCG 714

Qy 222 GGCCTGCTGCTGCTACTTCGCGGTTCGAACTCTCTCGGGAGCGCCGACGCGCTTGAGCG 281  
 |||  
 Db 715 CGCGGCG 774

Qy 282 GCGGGAAGACGACCTTCGCGGAGTCTTCGCGCGCGGTGCGTTCGCGGTTCAAGGAGCG 341  
 |||  
 Db 775 GCG 834

Qy 342 GGGCGTGTGTGACACGCTGTCTCGGCTTCGCGGACACCGGACCGGACCGGACCGGACG 401  
 |||  
 Db 835 GGGCGGCG 893











GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 Seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgttctacagtcgggcac.....ccccgcgcagggctcgtgc 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15

Total number of hits satisfying chosen parameters: 13643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	21	4.2	303 10	BF896427
2	21	4.2	316 10	BF896426
3	20	4.0	546 28	BH516834
4	20	4.0	709 28	BH585470

19	3.8	115	9	AA065382
19	3.8	249	9	AA492813
19	3.8	324	9	AA881756
19	3.8	378	10	BE418061
19	3.8	420	12	BG837626
19	3.8	431	10	BF411078
19	3.8	438	14	CB288581
19	3.8	445	9	A1746181
19	3.8	446	9	A1833600
19	3.8	449	14	CA402106
19	3.8	469	12	BG267470
19	3.8	474	14	CF608321
19	3.8	476	14	CF608208
19	3.8	485	12	BG265809
19	3.8	507	14	CB863203
19	3.8	522	9	A1737095
19	3.8	528	9	A1737056
19	3.8	530	14	CA401348
19	3.8	538	14	CF486345
19	3.8	541	9	A1677352
19	3.8	543	14	CA404459
19	3.8	544	14	CF488024
19	3.8	547	14	CB288531
19	3.8	562	14	CB288483
19	3.8	567	14	CA398780
19	3.8	571	9	A1714654
19	3.8	578	28	CC026191
19	3.8	581	14	CF489087
19	3.8	583	14	CB288474
19	3.8	584	12	BG266929
19	3.8	586	9	A1711641
19	3.8	590	9	A1657469
19	3.8	601	14	CD444963
19	3.8	610	9	A1759026
19	3.8	616	14	CB919676
19	3.8	617	9	A1745808
19	3.8	617	9	AI833967
19	3.8	621	12	BM356065
19	3.8	625	14	CD428301
19	3.8	631	14	CD442989
19	3.8	636	14	CA400622

ALIGNMENTS

RESULT 1

BF896427 303 bp mRNA linear EST 18-JAN-2001  
LOCUS RCL-WT0147-241100-011-f10 MT0147 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF896427  
ACCESSION BF896427.1 GI:12287886  
VERSION BF896427.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 303)  
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800

COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,



Thu Jun 17 09:02:23 2004

```

RESULT 4
BH585470/c
LOCUS
DEFINITION
  BH585470 BOGE Brassica oleracea genomic clone BOGEK50, genomic
  survey sequence.
ACCESSION
  BH585470
VERSION
  BH585470.1 GI:117837928
KEYWORDS
  GSS.
SOURCE
  Brassica oleracea
  Brassica oleracea
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
  1 (bases 1 to 709)
  Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
  Whole genome shotgun sequencing of Brassica oleracea
  Unpublished (2001)
  Other GSSs: BOGEK50TF
  Contact: Chris Town
TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA.
  Tel: 301-838-3523
  Fax: 301-838-0208
  Email: cdtown@tigr.org
  DNA is from a doubled haploid provided by Tom Osborn.
  Seq primer: TR
  Class: sheared ends.
FEATURES
  source
    1..709
    /organism="Brassica oleracea"
    /mol_type="genomic DNA"
    /strain="TO1000DH3"
    /db_xref="taxon:3712"
    /clone="BOGEK50"
    /clone_lib="BOGE"
    /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
    genomic DNA inserted into PHOS1 using BstXI linkers"
ORIGIN
  Query Match 4.0%; Score 20; DB 28; Length 709;
  Best Local Similarity 100.0%; Pred. No. 3.5e+02;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 226 GACGACGTGATCGGCTGAT 245
  Db 431 GACGACGTGATCGGCTGAT 412
  ||||||||||||||||||
RESULT 5
AA065382
LOCUS
DEFINITION
  m151e05.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
  clone IMAGE:515552 5', mRNA sequence.
ACCESSION
  AA065382.1 GI:1562663
VERSION
  AA065382
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 115)
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
  Waterston, R.
  The WashU-HHMI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HHMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:530607
  Seq primer: -28ml3 rev1 ET from Amersham.
FEATURES
  source
    1..115
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="CD-1"
    /db_xref="taxon:10090"
    /clone="IMAGE:515552"
    /sex="males"
    /tissue_type="testis"
    /dev_stage="10-12 week old"
    /lab_host="SOLR (kanamycin resistant)"
    /clone_lib="Stratagene mouse testis (#937308)"
    /note="Organ: testis; Vector: pBluescript SK-; Site 1:
    EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
    Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
    -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor
    sequence: 5' CTCGAGTGTGTGTGTGTGT 3'"
ORIGIN
  Query Match 3.8%; Score 19; DB 9; Length 115;
  Best Local Similarity 100.0%; Pred. No. 8.6e+02;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 267 CTGCGCGCGCGCGCGAC 285
  Db 21 CTGCGCGCGCGCGCGAC 39
  ||||||||||||||||||
RESULT 6
AA492813
LOCUS
DEFINITION
  vi78h04.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
  clone IMAGE:918391 5', mRNA sequence.
ACCESSION
  AA492813
VERSION
  AA492813.1 GI:2222375
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 249)
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
  Waterston, R.
  The WashU-HHMI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HHMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:530607
  Seq primer: -28ml3 rev1 ET from Amersham.
FEATURES
  source
    1..249
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="CD-1"
    /db_xref="taxon:10090"
  
```

```

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:309400
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 88.
FEATURES
  source
    1..115
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="CD-1"
    /db_xref="taxon:10090"
    /clone="IMAGE:515552"
    /sex="males"
    /tissue_type="testis"
    /dev_stage="10-12 week old"
    /lab_host="SOLR (kanamycin resistant)"
    /clone_lib="Stratagene mouse testis (#937308)"
    /note="Organ: testis; Vector: pBluescript SK-; Site 1:
    EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
    Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
    -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor
    sequence: 5' CTCGAGTGTGTGTGTGT 3'"
ORIGIN
  Query Match 3.8%; Score 19; DB 9; Length 115;
  Best Local Similarity 100.0%; Pred. No. 8.6e+02;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 267 CTGCGCGCGCGCGCGAC 285
  Db 21 CTGCGCGCGCGCGCGAC 39
  ||||||||||||||||||

```

```

RESULT 6
AA492813
LOCUS
DEFINITION
  vi78h04.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
  clone IMAGE:918391 5', mRNA sequence.
ACCESSION
  AA492813
VERSION
  AA492813.1 GI:2222375
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 249)
  Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
  Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
  Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
  Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
  Waterston, R.
  The WashU-HHMI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HHMI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:530607
  Seq primer: -28ml3 rev1 ET from Amersham.
FEATURES
  source
    1..249
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="CD-1"
    /db_xref="taxon:10090"
  
```

Thu Jun 17 09:02:23 2004

```

/clone="IMAGE:918391"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse testis (#937308)"
/notes="Organ: testis; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 249;
Best Local Similarity 100.0%; Pred. No. 9.1e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 267 CCTGGCGCGGTCCGACC 285
Db 144 CCTGGCGCGGTCCGACC 162

RESULT 7
AA881756 324 bp mRNA linear EST 26-MAR-1998
LOCUS vx27b02.r1 Soares mammary_gland NbMWG Mus musculus cDNA clone
DEFINITION IMAGE:1265643 5' similar to SW:FNPP_SCHPO Q00472
4-NITROPHENYLPHOSPHATASE 1, mRNA sequence.
ACCESSION AA881756
VERSION AA881756.1 GI:2991066
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 324)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:668195
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 308.
Location/Qualifiers
1..324
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1265643"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary gland NbMWG"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGCGCGCGGCGGAATGTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI

```

```

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 9.3e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 267 CCTGGCGCGGTCCGACC 285
Db 230 CCTGGCGCGGTCCGACC 248

RESULT 8
BE418061/c 378 bp mRNA linear EST 24-JUL-2000
LOCUS SCL011.H02R990126 ITEC SCL Wheat Leaf Library Triticum aestivum
DEFINITION cDNA clone SCL011.H02, mRNA sequence.
ACCESSION BE418061
VERSION BE418061.1 GI:9415809
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 378)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Sharifou,M., Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
JOURNAL Unpublished (2000)
COMMENT Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1..378
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL011.H02"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/clone_lib="ITEC SCL Wheat Leaf Library"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."

ORIGIN
Query Match 3.8%; Score 19; DB 10; Length 378;
Best Local Similarity 100.0%; Pred. No. 9.4e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 205 ATCGCGCGCGGCGGCTCCG 223
Db 258 ATCGCGCGCGGCGGCTCCG 240

RESULT 9
BG837626/c 420 bp mRNA linear EST 25-MAY-2001
LOCUS BG837626
DEFINITION Zm10_04b04_A Zm10_AAFc_EORC_Fusarium graminearum_corn_silk Zea

```

Thu Jun 17 09:02:23 2004

MEDLINE  
 PUBMED  
 COMMENT

97044477  
 889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Oligo-dt track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA library Preparation: M.B.  
 Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=No.

FEATURES  
 Location/Qualifiers  
 1..431  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CNO-bmg-e-12-0-UI"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-CNO"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-CNO  
 library is a normalized library constructed from the  
 following tissues: brown adipose, penis, salivary glands,  
 bladder, fundus, cervix, seminal vesicles. For a detailed  
 description of the library from which this clone was  
 derived, please visit our web site at  
 ratseq.eng.uiowa.edu. The subtraction has been previously  
 described in (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)  
 TAG\_SEQ=None found"

ORIGIN  
 Query Match 3.8%; Score 19; DB 10; Length 431;  
 Best Local Similarity 100.0%; Pred.No. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 360 GGGCGCGGTCTCCGGCGCC 378  
 |||||  
 Db 335 GGGCGCGGTCTCCGGCGCC 353

RESULT 11  
 CB288581/C  
 LOCUS  
 DEFINITION  
 mRNA sequence.  
 V-B-16F09 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-16F09 5',  
 CB288581  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Vitis aestivalis  
 Vitis aestivalis  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; Vitaceae; Vitis.  
 1 (bases 1 to 438)  
 Hou,H.S., Phanikant,T.V., Kovacs,L. and Qiu,W.P.  
 Expressed sequence tags of young leaf tissues of a  
 disease-resistant Vitis aestivalis var. Norton  
 Unpublished (2003)  
 JOURNAL  
 COMMENT  
 Contact: Wenping Qiu  
 Department of Fruit Science  
 Southwestern Missouri State University-Mountain Grove  
 9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA  
 Tel: 417 926 4105  
 Fax: 417 926 6646  
 Email: weq070f@msu.edu  
 Inherit Length: 438 Std Error: 0.00

mays cDNA clone Zm10\_04b04, mRNA sequence.  
 BG537626  
 VERSION  
 BG537626.1 GI:14203949  
 KEYWORDS  
 EST.  
 ORGANISM  
 Zea mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 420)  
 Harris,L.J., Balcerak,M., Allard,S., Saparno,A., Couroux,P., De  
 Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A.,  
 Spott,D. and Tinker,N.A.  
 Expressed Sequence Tags from Maize Silk Six Hours After Silk  
 Channel Inoculation with Fusarium graminearum  
 Unpublished (2001)  
 JOURNAL  
 COMMENT  
 Contact: Harris, Linda J.  
 Eastern Cereal and Oilseed Research Centre  
 Agriculture and Agri-food Canada  
 Bldg. 21, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,  
 CANADA  
 Tel: (613) 759-1314  
 Fax: (613) 759-6566  
 Email: harrislj@em.agr.ca.  
 Location/Qualifiers  
 1..420  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="CO388"  
 /db\_xref="taxon:4577"  
 /clone="Zm10\_04b04"  
 /tissue\_type="Silk"  
 /dev\_stage="4-5 days post-silk emergence"  
 /clone\_lib="Zm10\_AAFCECFC\_Fusarium\_graminearum\_corn\_silk"  
 /note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;  
 Site 2: XhoI; Field-grown corn was silk channel-inoculated  
 in the morning (~10 am) with 1 ml of a macroconidial  
 suspension (500,000 spores/ml) of Fusarium graminearum and  
 silk channels were collected and immediately frozen in  
 liquid nitrogen 6 hours later. RNA was extracted from  
 silk tissue between 1 cm below and above the inoculation  
 point in the silk channel, RNA from five silk channels was  
 pooled."

ORIGIN  
 Query Match 3.8%; Score 19; DB 12; Length 420;  
 Best Local Similarity 100.0%; Pred.No. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 125 TCTGGCGGTCTCCGGCTGT 143  
 |||||  
 Db 204 TCTGGCGGTCTCCGGCTGT 186

RESULT 10  
 BF411078  
 LOCUS  
 DEFINITION  
 UI-R-CNO-bmg-e-12-0-UI.s1 UI-R-CNO Rattus norvegicus cDNA clone  
 UI-R-CNO-bmg-e-12-0-UI 3', mRNA sequence.  
 BF411078  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 431)  
 Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)

Thu Jun 17 09:02:23 2004

Plate: VAN-Baker-1-6 row: F column: 09  
 Seq primer: T3 PRIMER  
 High quality sequence stop: 438  
 POLYA=NO.

## FEATURES

source

1. .438  
 /organism="Vitis aestivalis"  
 /mol\_type="mRNA"  
 /cultivar="Norton"  
 /db\_xref="taxon:3605"  
 /clone="V-B-16F09"  
 /tissue\_type="leaf"  
 /dev\_stage="Young leaf"  
 /lab\_host="Xl10-Gold E. coli"  
 /clone\_lib="VAN-Baker-1"  
 /note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho I; Site 2: EcoRI; VAN-Baker-1 is a cDNA library of Norton grape young leaves (Vitis aestivalis var. Norton) grapevines were grown under normal greenhouse conditions. The cDNA synthesis and library construction was performed according to the instruction manual for pBluescript II XR cDNA library construction kit provided by Stratagene."

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 221 CGCCGACGACGTGATGCG 239

Db 140 CGCCGACGACGTGATGCG 122

## RESULT 12

AI746181 445 bp mRNA linear EST 02-FEB-2000  
 LOCUS 605082C07.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays  
 DEFINITION cDNA, mRNA sequence.

ACCESSION AI746181.1 GI:5124445

VERSION AI746181

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 445)

REFERENCE

AUTHORS

TITLE

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605082 row: C column: 07.

Location/Qualifiers

1. .445

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="Ohio43"

/db\_xref="taxon:4577"

/tissue\_type="nucellar, embryo, and endospERM"

/dev\_stage="10-14 days post-pollination"

/lab\_host="DH5(alpha)"

/clone\_lib="605 - EndospERM cDNA library from Schmidt lab"

/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI; Site 2: XhoI; Kernel endospERM cDNA library from Schmidt lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 271 GCCGCGGTGCCGACCACCT 289

Db 366 GCCGCGGTGCCGACCACCT 384

## RESULT 13

AI833600

LOCUS

DEFINITION

605092F11.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays

cDNA, mRNA sequence.

ACCESSION AI833600

VERSION AI833600.1 GI:5455910

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 446)

REFERENCE

AUTHORS

TITLE

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605092 row: F column: 11.

Location/Qualifiers

1. .446

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="Ohio43"

/db\_xref="taxon:4577"

/tissue\_type="nucellar, embryo, and endospERM"

/dev\_stage="10-14 days post-pollination"

/lab\_host="DH5(alpha)"

/clone\_lib="605 - EndospERM cDNA library from Schmidt lab"

/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI; Site 2: XhoI; Kernel endospERM cDNA library from Schmidt lab"

lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 446;  
 Best Local Similarity 100.0%; Pred. No. 9.5e+02; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

QY 271 GCCGCGGTGCCGACCACCT 289

Db 374 GCCGCGGTGCCGACCACCT 392

## RESULT 14

CA402106

LOCUS

DEFINITION

E01N0431E02.9 EndospERM\_4 Zea mays cDNA, mRNA sequence.

ACCESSION CA402106

VERSION CA402106.1 GI:24766961

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 449)

REFERENCE



Thu Jun 17 09:02:23 2004

AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and Messing, J.

TITLE Sequencing of the maize endosperm ESTs

JOURNAL Unpublished (2002)

COMMENT Contact: Lai, Jinsheng

Dr. Joachim Messing's lab

Waksman Institute, Rutgers University

190 Frelinghuysen Rd., Piscataway, NJ 08854, USA

Tel: 732-445-3801

Fax: 732-445-5735

Email: jlai@waksman.rutgers.edu

Seq primer: T7.

FEATURES

source

1. .449 Location/Qualifiers

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="W22"

/db\_xref="taxon:4577"

/tissue\_type="Endosperm of 7-23DAP"

/clone\_lib="Endosperm\_4"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

XhoI

ORIGIN

Query Match

Best Local Similarity 3.8%; Score 19; DB 14; Length 449;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289

Db 411 GCCGGCGTGGCGACCACT 429

RESULT 15

LOCUS BG267470

DEFINITION 1000122C06.x1 1000 - Unigene 1 from Maize Genome Project Zea mays

ACCSSION BG267470

VERSION BG267470.1 GI:12971414

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 469)

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 1000122 row: C column: 06.

Location/Qualifiers

1. .469

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="dbEST:605097G06.x1"

/db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene 1 from Maize Genome Project"

/note="This library represents the unique ESTs found in

for the first round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

libraries 486, 487, 496, 603, 605, 606, 614, 618, 660,

683, 687, 707, and 945. Contigs were assembled using

TIGR's CAP program and a representative EST from each

contig was selected for the Unigene set. All singlets were

ORIGIN

Query Match

Best Local Similarity 3.8%; Score 19; DB 12; Length 469;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289

Db 391 GCCGGCGTGGCGACCACT 409

RESULT 16

LOCUS CF608321/c

DEFINITION GENMA01\_001873 Grape Bud pSPORT1 Library Vitis vinifera cDNA 5',

ACCSSION CF608321

VERSION CF608321.1 GI:37188972

KEYWORDS EST.

SOURCE Vitis vinifera

ORGANISM Vitis vinifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosids; Vitaceae; Vitis.

1 (bases 1 to 474)

Moser, C., Segala, C., Fontana, P., Salakhudinov, I., Gatto, P.,

Pindo, M., Zyprian, E., Toepfer, R., Grando, M.S. and Velasco, R.

Expressed sequence tags from different organs of Vitis vinifera

Unpublished (2003)

Contact: Moser C

Laboratorio di Genetica Molecolare

Istituto Agrario di San Michele all'Adige (IASMA)

via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia

Tel: 0039-0461-615314

Fax: 0039-0461-650956

Email: claudio.moser@ismaa.it

The sequencing work has been funded by the 'Fondazione Cassa di

Risparmio di Trento e Rovereto'

High quality sequence stop: 474.

Location/Qualifiers

1. .474

/organism="Vitis vinifera"

/mol\_type="mRNA"

/cultivar="Pinot noir"

/db\_xref="taxon:29760"

/sex="Hermaphrodite"

/dev\_stage="bud swelling"

/lab\_host="DH10B"

/clone\_lib="Grape Bud pSPORT1 Library"

/note="Organ: bud; Vector: pSPORT1; Site\_1: NctI; Site\_2:

Sall"

ORIGIN

Query Match

Best Local Similarity 3.8%; Score 19; DB 14; Length 474;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGACGATGCG 239

Db 147 CCGCCGACGACGATGCG 129

RESULT 17

LOCUS CF608208/c

DEFINITION GENMA01\_001751 Grape Bud pSPORT1 Library Vitis vinifera cDNA 5',

ACCSSION CF608208

VERSION CF608208.1 GI:37188859

KEYWORDS EST.

SOURCE Vitis vinifera

ORGANISM Vitis vinifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 476)  
Moser, C., Segala, C., Fontana, P., Salakhutdinov, I., Gatto, P., Pindo, M., Zyprian, E., Toepfer, R., Grando, M. S. and Velasco, R.  
Expressed sequence tags from different organs of Vitis vinifera  
Unpublished (2003)  
Contact: Moser C  
Laboratorio di Genetica Molecolare  
Istituto Agrario di San Michele all'Adige (IASMA)  
Via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia  
Tel: 0039-0461-615314  
Fax: 0039-0461-650956

Email: claudio.moser@ismaa.it  
The sequencing work has been funded by the 'Fondazione Cassa di Risparmio di Trento e Rovereto'  
High quality sequence stop: 476.  
Location/Qualifiers

## FEATURES

source

1..476  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/cultivar="Pinot noir"  
/db\_xref="taxon:29760"  
/sex="Hermaphrodite"  
/dev\_stage="bud swelling"  
/lab\_host="DH10B"  
/clone\_lib="Grape Bud pSPORT1 Library"  
/note="Organ: bud; Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 476;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCGGACGCGTATCGC 239

Db 145 CCGCGGACGCGTATCGC 127

## RESULT 18

## BG266809

## LOCUS

1000106A07.x1 1000 - Unigene I from Maize Genome Project Zea mays  
CDNA, mRNA sequence.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 1000106 row: A column: 07.

## FEATURES

source

1..485  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="dbEST:605025C02.x1"  
/db\_xref="taxon:4577"

Query Match 3.8%; Score 19; DB 14; Length 507;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 GCCGGACGGCGCCCGGTC 489

Db 328 GCCGGACGGCGCCCGGTC 346

## ORIGIN

Query Match 3.8%; Score 19; DB 12; Length 485;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 GCCGGACGGCGCCCGGTC 489

Db 328 GCCGGACGGCGCCCGGTC 346

/clone\_lib="1000 - Unigene I from Maize Genome Project"  
/note="This library represents the unique ESTs found in the first round of EST sequencing at Stanford University for the maize genome project. Sequences are present from libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683, 687, 707, and 945. Contigs were assembled using TIGR's CAP program and a representative EST from each contig was selected for the Unigene set. All singlets were also selected."

## ORIGIN

Query Match 3.8%; Score 19; DB 12; Length 485;

Best Local Similarity 100.0%; Pred. No. 9.6e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289

Db 412 GCCGGCGTCCGACCACT 430

## RESULT 19

## CB863203

## LOCUS

HH04A07w HH Hordeum vulgare cDNA clone HH04A07 5-PRIME, mRNA  
sequence.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.  
1 (bases 1 to 507)  
Varshney, R.K., Zhang, H., Burton, R., Stein, N., Langridge, P. and Graner, A.  
Barley ESTs from coleoptile tissue

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished (2003)  
Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert Length: 507  
Plate: 4 row: A column: 7  
Seq primer: T7.

## FEATURES

source

1..507  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/cultivar="Sloop"  
/db\_xref="GABI:554860"  
/db\_xref="taxon:4513"  
/clone="HH04A07"  
/tissue\_type="coleoptile"  
/dev\_stage="coleoptile, 1 day old"  
/lab\_host="DH10B"  
/clone\_lib="HH"  
/note="Vector: pSPORT; Site\_1: SalI (5-end of cDNA); Site\_2: NotI (3-end of cDNA); Due to the cloning system used blue/white selection for recombinants is not 100% reliable. Average insert size is 1.3 kb."

Thu Jun 17 09:02:23 2004

us-10-042-665a-3\_copy\_1825\_2325.ol.rst

```

RESULT 20
LOCUS      AI737095          522 bp    mRNA    linear    EST 02-FEB-2000
DEFINITION 605017Fl1.x2 605 - EndospERM cDNA library from Schmidt lab Zea mays
ACCESSION  AI737095
VERSION     AI737095.1  GI:5058619
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 522)
AUTHORS    Walbot V.
TITLE      Zea mays from various cDNA libraries sequenced at Stanford
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605017 row: F column: 11.

FEATURES   source
            Location/Qualifiers
            1..522
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /tissue_type="mucellar, embryo, and endospERM"
            /lab_host="DHS(alpha)"
            /dev_stage="10-14 days post-pollination"
            /clone_lib="605 - EndospERM cDNA library from Schmidt lab"
            /note="Organ: Kernel; Vector: PAD-GAL4-2; Site_1: EcoRI;
            Site_2: XhoI; Kernel endospERM cDNA library from Schmidt
            lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 522;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
      |||||
DB 418 GCCGGCGTGGCGACCACT 436

RESULT 21
LOCUS      AI737056          528 bp    mRNA    linear    EST 02-FEB-2000
DEFINITION 605017B10.x2 605 - EndospERM cDNA library from Schmidt lab Zea mays
ACCESSION  AI737056
VERSION     AI737056.1  GI:5058580
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 528)
AUTHORS    Walbot V.
TITLE      Zea mays from various cDNA libraries sequenced at Stanford
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University

FEATURES   source
            Location/Qualifiers
            1..522
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="Ohio43"
            /db_xref="taxon:4577"
            /tissue_type="mucellar, embryo, and endospERM"
            /lab_host="DHS(alpha)"
            /dev_stage="10-14 days post-pollination"
            /clone_lib="605 - EndospERM cDNA library from Schmidt lab"
            /note="Organ: Kernel; Vector: PAD-GAL4-2; Site_1: EcoRI;
            Site_2: XhoI; Kernel endospERM cDNA library from Schmidt
            lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 522;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
      |||||
DB 418 GCCGGCGTGGCGACCACT 436

RESULT 22
LOCUS      CA401348          530 bp    mRNA    linear    EST 07-NOV-2002
DEFINITION EL01N0419A06.g EndospERM_4 Zea mays cDNA, mRNA sequence.
ACCESSION  CA401348
VERSION     CA401348.1  GI:24766193
KEYWORDS   Zea mays
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 530)
AUTHORS    Lai J., Dey N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
            Messing, J.
TITLE      Sequencing of the maize endospERM ESTs
JOURNAL    Unpublished (2002)
COMMENT    Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Wakeman Institute, Rutgers University
            190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
            Tel: 732-445-3801
            Fax: 732-445-5735
            Email: jlai@waksman.rutgers.edu
            Seq primer: T7.

FEATURES   source
            Location/Qualifiers
            1..530
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="W22"
            /db_xref="taxon:4577"
            /tissue_type="EndospERM of 7-23DAP"
            /clone_lib="EndospERM_4"
            /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
            XhoI"

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
      |||||
DB 441 GCCGGCGTGGCGACCACT 459

```

855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 605017 row: B column: 10.

## FEATURES

source  
 Location/Qualifiers  
 1..528  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="Ohio43"  
 /db\_xref="taxon:4577"  
 /tissue\_type="mucellar, embryo, and endospERM"  
 /dev\_stage="10-14 days post-pollination"  
 /lab\_host="DHS(alpha)"  
 /clone\_lib="605 - EndospERM cDNA library from Schmidt lab"  
 /note="Organ: Kernel; Vector: PAD-GAL4-2; Site\_1: EcoRI;  
 Site\_2: XhoI; Kernel endospERM cDNA library from Schmidt  
 lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 271 GCCGGCGTGGCGACCACT 289  
 |||||  
 DB 409 GCCGGCGTGGCGACCACT 427

## RESULT 22

LOCUS CA401348 530 bp mRNA linear EST 07-NOV-2002  
 DEFINITION EL01N0419A06.g EndospERM\_4 Zea mays cDNA, mRNA sequence.  
 ACCESSION CA401348  
 VERSION CA401348.1 GI:24766193  
 KEYWORDS Zea mays  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 530)  
 AUTHORS Lai J., Dey N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and  
 Messing, J.  
 TITLE Sequencing of the maize endospERM ESTs  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Lai, Jinsheng  
 Dr. Joachim Messing's lab  
 Wakeman Institute, Rutgers University  
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
 Tel: 732-445-3801  
 Fax: 732-445-5735  
 Email: jlai@waksman.rutgers.edu  
 Seq primer: T7.

## FEATURES

source  
 Location/Qualifiers  
 1..530  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="W22"  
 /db\_xref="taxon:4577"  
 /tissue\_type="EndospERM of 7-23DAP"  
 /clone\_lib="EndospERM\_4"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 271 GCCGGCGTGGCGACCACT 289  
 |||||  
 DB 441 GCCGGCGTGGCGACCACT 459

```

RESULT 23
CF486345/c
LOCUS
DEFINITION
POL1_37_E02_b1_A002 Pollen Sorghum bicolor cDNA clone
EST 08-SEP-2003
CF486345
VERSION
KEYWORDS
SOURCE
ORGANISM
Sorghum bicolor (sorghum)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 538)
Cordonnier-Pratt M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,
Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,
Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and
Pratt, D.H.
EST database from Sorghum: pollen
Unpublished (2003)
Other ESTs: POL1_37_E02_g1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug3 (CGACCTGCGACTCGAGCACA)
POLYA=No.

FEATURES
source
1..538
Location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="BX623"
/clone="POL1_37_E02_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
Site 2: XhoI; The library was prepared from polyA+ RNA
from pollen at the late vacuolated-vacuolated stage of
development. Pollen was harvested from greenhouse-grown
panicles of sorghum line BTx623. Panicles were removed
from the flag leaf prior to emergence, when no detectable
amylase is present in pollen of male-fertile lines. This
stage represents pollen collected from anthers about 8-14
days prior to anthesis. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
insert."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 538;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
Db 129 GCCGGCGTGGCGACCACT 111

RESULT 24
AI677252
LOCUS
DEFINITION
605052A08.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
AI677252
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 541)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605052 row: A column: 08.
Location/Qualifiers
1..541
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="pH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site 1: EcoRI;
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 541;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTGGCGACCACT 289
Db 391 GCCGGCGTGGCGACCACT 409

RESULT 25
CA401459
LOCUS
DEFINITION
E01N0421B10.9 Endosperm_4 Zea mays cDNA, mRNA sequence.
CA401459
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 543)
Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
Sequencing of the maize endosperm ESTs
Unpublished (2002)
Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735

```

Email: jlai@waksman.rutgers.edu

Seq primer: 17

#### FEATURES

source

Location/Qualifiers  
1..543  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 543;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289

DB 455 GCCGGCGTCCGACCACT 473

#### RESULT 26

CF488024/c

LOCUS

DEFINITION

POLL\_47 G09 B1 A002 Pollen Sorghum bicolor cDNA clone

CF488024

CF488024.1 GI:34516893

EST.

Sorghum bicolor (sorghum)

ORGANISM

KEYWORDS

Sorghum bicolor

REFERENCE

AUTHORS

1 (bases 1 to 544)

Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,

Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J.,

Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and

Pratt, L.H.

EST database from Sorghum: pollen

Unpublished (2003)

Other ESTs: POLL\_47 G09 gl A002

Contact: Cordonnier-Pratt M

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A & M University;

sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences

are presented as their reverse complement and have been trimmed to

exclude polyA.

Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC)

POLYA=Yes.

Location/Qualifiers

1..544

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone="POLL\_47 G09 A002"

/lab\_host="DH10B-T1 phage-resistant E. coli"

/clone\_lib="Pollen"

/note="Organ: Pollen; Vector: pME18S-FL3; Site\_1: XhoI;

Site\_2: XhoI; The library was prepared from polyA+ RNA

from pollen at the late vacuolated-vacuolated stage of

development. Pollen was harvested from greenhouse-grown panicles of sorghum line Brx623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 544;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289

DB 129 GCCGGCGTCCGACCACT 111

#### RESULT 27

CB288531/c

LOCUS

DEFINITION

V-B-16B03 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-16B03 5',

mRNA sequence.

ACCESSION

CB288531

VERSION

CB288531.1 GI:28602272

KEYWORDS

EST.

SOURCE

Vitis aestivalis

ORGANISM

Vitis aestivalis

REFERENCE

1 (bases 1 to 547)

Hou, H.S., Phamkhanh, T.V., Kovacs, L., and Qiu, W.P.

Expressed sequence tags of young leaf tissues of a

disease-resistant Vitis aestivalis var. Norton

Unpublished (2003)

CONTACT: Wenping Qiu

Department of Fruit Science

Southwest Missouri State University-Mountain Grove

9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA

Tel: 417 926 4105

Fax: 417 926 6646

Email: weq070f@msu.edu

Insert Length: 547 Std Error: 0.00

Plate: VAN-Baker-1-6 row: B column: 03

Seq primer: T3 PRIMER

High quality sequence stop: 547

POLYA=No.

Location/Qualifiers

1..547

/organism="Vitis aestivalis"

/mol\_type="mRNA"

/cultivar="Norton"

/db\_xref="taxon:3605"

/clone="V-B-16B03"

/tissue\_type="Leaf"

/dev\_stage="Young leaf"

/lab\_host="X1010-Gold E. coli"

/clone\_lib="VAN-Baker-1"

/note="Vector: pBluescript II SK (+) Phagemid; Site\_1: Xho

I; Site\_2: EcoR I; VAN-Baker-1 is a cDNA library of Norton

grape young leaves (Vitis aestivalis var. Norton). Norton

grapevines were grown under normal greenhouse conditions.

The cDNA synthesis and library construction was performed

according to the instruction manual for pBluescript II XR

cDNA library construction kit provided by Stratagene."

#### ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 547;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;

```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGCGTGATGCG 239
      |||||
Db 114 CCGCCGACGCGTGATGCG 96

RESULT 28
CB288483/c
LOCUS V-B-15E08 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-15E08 5',
DEFINITION mRNA sequence.
ACCESSION CB288483
VERSION CB288483.1 GI:28602224
KEYWORDS EST.
SOURCE Vitis aestivalis
ORGANISM Vitis aestivalis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 562)
AUTHORS Hou,H.S.; Phanikanth,T.V.; Kovacs,L. and Qiu,W.P.
TITLE Expressed sequence tags of young leaf tissues of a
disease-resistant Vitis aestivalis var. Norton
JOURNAL Unpublished (2003)
COMMENT Contact: Wenping Qiu
Department of Fruit Science
Southwest Missouri State University-Mountain Grove
9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA
Tel: 417 926 4105
Fax: 417 926 6646
Email: weq070f@msu.edu
Insert Length: 562 Std Error: 0.00
Plate: VAN-Baker-1-5 row: E column: 08
Seq primer: T3 PRIMER
High quality sequence stop: 562
POLYA=No.

FEATURES             Location/Qualifiers
     source
     1..562
     /organism="Vitis aestivalis"
     /mol_type="mRNA"
     /cultivar="Norton"
     /db_xref="taxon:3605"
     /clone="V-B-15E08"
     /tissue_type="leaf"
     /dev_stage="young leaf"
     /lab_host="XL10-Gold E .coli"
     /clone_lib="VAN-Baker-1"
     /note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho
I; Site 2: EcoR I; VAN-Baker-1 is a cDNA library of Norton
grape young leaves (Vitis aestivalis var. Norton). Norton
grapevines were grown under normal greenhouse conditions.
The cDNA synthesis and library construction was performed
according to the instruction manual for pBluescript II XR
cDNA library construction kit provided by Stratagene."

ORIGIN
Query Match          3.8%; Score 19; DB 14; Length 562;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGCGTGCCGACCACCT 289
      |||||
Db 504 GCCGCGGTGCCGACCACCT 522

RESULT 30
AI714654
LOCUS 605068G06.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI714654
VERSION AI714654.1 GI:5018461
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 571)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605068 row: G column: 06.

FEATURES             Location/Qualifiers
     source
     1..571
     /organism="Zea mays"
     /mol_type="mRNA"
     /cultivar="Ohio43"
     /db_xref="taxon:4577"
     /tissue_type="nucellar, embryo, and endospERM"
     /dev_stage="10-14 days post-pollination"

QY 221 CCGCCGACGCGTGATGCG 239
      |||||
Db 142 CCGCCGACGCGTGATGCG 124

RESULT 29
CA398780
LOCUS CA398780
DEFINITION EL01N0309H09.g EndospERM_3 Zea mays cDNA, mRNA sequence.
ACCESSION CA398780
VERSION CA398780.1 GI:24763608
KEYWORDS EST.

```

```

SOURCE
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 567)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endospERM ESTs
JOURNAL Unpublished (2002)
COMMENT Contact: Lai, jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T7.
FEATURES             Location/Qualifiers
     source
     1..567
     /organism="Zea mays"
     /mol_type="mRNA"
     /cultivar="W22"
     /db_xref="taxon:4577"
     /tissue_type="EndospERM of 7-23DAP"
     /clone_lib="EndospERM_3"
     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match          3.8%; Score 19; DB 14; Length 567;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGCGGTGCCGACCACCT 289
      |||||
Db 504 GCCGCGGTGCCGACCACCT 522

RESULT 30
AI714654
LOCUS 605068G06.xl 605 - EndospERM cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI714654
VERSION AI714654.1 GI:5018461
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 571)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605068 row: G column: 06.

FEATURES             Location/Qualifiers
     source
     1..571
     /organism="Zea mays"
     /mol_type="mRNA"
     /cultivar="Ohio43"
     /db_xref="taxon:4577"
     /tissue_type="nucellar, embryo, and endospERM"
     /dev_stage="10-14 days post-pollination"

```

/lab host="DH5 (alpha)"  
 /clone lib="605 - Endospore cDNA library from Schmidt lab"  
 /note="Organ: Kernel; Vector: PAD-GAL4-2"; Site 1: EcoRI;  
 Site 2: XhoI; Kernel endospore cDNA library from Schmidt  
 lab"

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 571;  
 Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCCACCACT 289

Db 471 GCCGGCGTGGCCACCACT 489

## RESULT 31

CC026191

LOCUS

3591.1.4.1.G04.x.1 3591 - RescueMu Grid P Zea mays genomic, genomic  
 survey sequence.

ACCESSION

CC026191

VERSION

GSS.

KEYWORDS

Zea mays

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 578)

Walbot.V.

Maize genomic sequences found using engineered RescueMu transposon

Unpublished (2001)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3591.1.4.1 row: 3

Class: transposon-tagged.

## FEATURES

source

1..578

/organism="Zea mays"

/mol\_type="genomic DNA"

/cultivar="mixed background W23/A188/B73/X55"

/db\_xref="taxon:4577"

/tissue\_type="leaf"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="3591 - RescueMu Grid P"

/note="Organ: leaf; Vector: RescueMu (engineered from

pBluescript backbone); Site 1: BamHI; Site 2: BglII;

RescueMu is a 4.9 kb, modified maize Mu transposon

designed to allow plasmid rescue from total genomic DNA.

Mu elements insert preferentially into transcription

units. For more information on RescueMu, go to the web

site 'www.zmdb.lasstate.edu' and follow the links for

'RescueMu.' Grid P was grown at Molokai in 2002. DNA was

extracted from leaf strips, double digested using BamHI

and BglII, and ligated to form circular plasmids. DH10B

cells were transformed and then screened on LB plates with

ampicillin."

## ORIGIN

Query Match 3.8%; Score 19; DB 28; Length 578;  
 Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 TCGCTGCTGCTGCTGCTG 179

Db 291 TCGCTGCTGCTGCTGCTG 309

## RESULT 32

CF489087/c

LOCUS

DEFINITION

POL1\_54\_G06.bl\_A002 Pollen Sorghum bicolor cDNA clone

POL1\_54\_G06\_A002 3', mRNA sequence.

ACCESSION

CF489087

VERSION

EST.

KEYWORDS

Sorghum bicolor (sorghum)

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 581)

Cordonnier-Pratt.M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,  
 Sun,F., Sullivan,R., Eastman,A. and Pratt,L.H.

An EST database from Sorghum: pollen

Unpublished (2003)

Other ESTs: POL1\_54\_G06.g1\_A002

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A & M University;

sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences

are presented as their reverse complement and have been trimmed to

exclude polyA.

Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACC)

POLYA=Yes.

## FEATURES

source

1..581

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone="POL1\_54\_G06\_A002"

/lab\_host="DH10B-T1 phage-resistant E. coli"

/clone\_lib="Pollen"

/note="Organ: Pollen; Vector: pME188-FL3; Site 1: XhoI;

Site 2: XhoI; The library was prepared from polyA+ RNA

from pollen at the late vacuolated-vacuolated stage of

development. Pollen was harvested from greenhouse-grown

panicles of sorghum line BTx623. Panicles were removed

from the flag leaf prior to emergence, when no detectable

amylase is present in pollen of male-fertile lines. This

stage represents pollen collected from anthers about 8-14

days prior to anthesis. Double-stranded cDNA was cloned

unidirectionally into different DraIII sites of the

pME188-FL3 vector (5-prime DraIII site is CACTGTGTG,

3-prime DraIII site is CACCAATGTG). XhoI excises the cDNA

insert."

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 581;

Best Local Similarity 100.0%; Pred. No. 9.7e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCCACCACT 289

Db 198 GCCGGCGTGGCCACCACT 180

## RESULT 33

CB288474/c

```

LOCUS      CB288474          583 bp      mRNA      linear      EST 27-FEB-2003
DEFINITION V-B-15D08 VAN-Baker-1 Vitis aestivalis cDNA clone V-B-15D08 5',
            mRNA sequence.
ACCESSION  CB288474
VERSION    CB288474.1      GI:28602215
KEYWORDS   EST.
SOURCE     Vitis aestivalis
ORGANISM   Vitis aestivalis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; Vitaceae; Vitis.
REFERENCE  1 (bases 1 to 583)
AUTHORS   Hou,H.S., Phanikant,T.V., Kovacs,L. and Qiu,W.P.
TITLE     Expressed sequence tags of young leaf tissues of a
            disease-resistant Vitis aestivalis var. Norton
JOURNAL    Unpublished (2003)
COMMENT    Contact: Wenping Qiu
            Department of Fruit Science
            Southwest Missouri State University-Mountain Grove
            9740 Red Spring Road, Mountain Grove, MO 65711-2999, USA
            Tel: 417 926 4105
            Fax: 417 926 6646
            Email: weq070f@msu.edu
            Insert length: 583 Std Error: 0.00
            Plate: VAN-Baker-1-5 row: D column: 08
            Seq primer: T3 PRIMER
            High quality sequence stop: 583
            POLYA=No.

FEATURES             Location/Qualifiers
     source
     1..583
     /organism="Vitis aestivalis"
     /mol_type="mRNA"
     /cultivar="Norton"
     /db_xref="taxon:3605"
     /clone="V-B-15D08"
     /tissue_type="Leaf"
     /dev_stage="young leaf"
     /lab_host="Xl10-Gold E .coli"
     /clone_lib="VAN-Baker-1"
     /note="Vector: pBluescript II SK (+) Phagemid; Site 1: Xho
I; Site 2: EcoR I; VAN-Baker-1 is a cDNA library of Norton
grape young leaves (Vitis. aestivalis var. Norton). Norton
grapevines were grown under normal greenhouse conditions.
The cDNA synthesis and library construction was performed
according to the instruction manual for pBluescript II XR
cDNA library construction kit provided by Stratagene."

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 583;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  221  CCGCGACGACGCGATGCG 239
Db      |||||||||||||||||||
        CCGCGACGACGCGATGCG 129

RESULT 34
LOCUS      BG266929          584 bp      mRNA      linear      EST 20-FEB-2001
DEFINITION 1000108H08.x2 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION  BG266929
VERSION    BG266929.1      GI:12970330
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 584)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

LOCUS      BG266929          584 bp      mRNA      linear      EST 02-FEB-2000
DEFINITION 605058F01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION  A1711641
VERSION    A1711641.1      GI:5005579
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 586)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

QY  271  GCCGGCGTGCCGACCACT 289
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACT 405

RESULT 35
LOCUS      A1711641          586 bp      mRNA      linear      EST 02-FEB-2000
DEFINITION 605058F01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION  A1711641
VERSION    A1711641.1      GI:5005579
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 586)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

QY  387  GCCGGCGTGCCGACCACT 405
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACT 405

ORIGIN
Query Match      3.8%; Score 19; DB 12; Length 584;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  271  GCCGGCGTGCCGACCACT 289
Db      |||||||||||||||||||
        GCCGGCGTGCCGACCACT 405

FEATURES             Location/Qualifiers
     source
     1..584
     /organism="Zea mays"
     /mol_type="mRNA"
     /db_xref="dbEST:605034D06.x1"
     /db_xref="taxon:4577"
     /clone_lib="1000 - Unigene I from Maize Genome Project"
     /note="This library represents the unique ESTs found in
the first round of EST sequencing at Stanford University
for the maize genome project. Sequences are present from
libraries 486, 487, 496, 603, 605, 614, 618, 660,
683, 687, 707, and 945. Contigs were assembled using
TIGR's CAP program and a representative EST from each
contig was selected for the Unigene set. All singlets were
also selected."

```



Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

ORIGIN  
Query Match 3.8%; Score 19; DB 9; Length 586;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289  
|||||  
Db 411 GCCGGCGTGGCGACCACT 429

RESULT 36  
Al657469 590 bp mRNA linear EST 02-FEB-2000  
LOCUS 605003F08.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
DEFINITION cDNA, mRNA sequence.

ACCESSION Al657469  
VERSION Al657469.1 GI:4753559  
KEYWORDS EST.

SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 590)

REFERENCE  
AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL  
COMMENT Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 605003 row: F column: 08.

FEATURES  
source Location/Qualifiers

1..590  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site 1: EcoRI;  
Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 590;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289  
|||||  
Db 393 GCCGGCGTGGCGACCACT 411

RESULT 37  
CD444963/c 601 bp mRNA linear EST 03-JUN-2003  
LOCUS EL01N0446A06.b Endosperm\_4 Zea mays cDNA, mRNA sequence.

DEFINITION CD444963  
ACCESSION CD444963  
VERSION CD444963.1 GI:31360606  
KEYWORDS EST.

SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and Messing,J.

TITLE Sequencing of the maize endosperm ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lai, Jinsheng

Dr. Joachim Messing's lab  
Waksman Institute, Rutgers University  
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu

Seq primer: T3.

FEATURES  
source Location/Qualifiers

1..601  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm 4"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI"

ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 601;  
Best Local Similarity 100.0%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTGGCGACCACT 289  
|||||  
Db 510 GCCGGCGTGGCGACCACT 492

RESULT 38

Al759026

LOCUS

DEFINITION Al759026 610 bp mRNA linear EST 02-FEB-2000  
605085H05.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays  
cDNA, mRNA sequence.

ACCESSION Al759026

VERSION Al759026.1 GI:5152861

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 610)

REFERENCE  
AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 605085 row: H column: 05.

FEATURES  
source Location/Qualifiers

1..610  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="Ohio43"  
/db\_xref="taxon:4577"  
/tissue\_type="nucellar, embryo, and endosperm"  
/dev\_stage="10-14 days post-pollination"  
/lab\_host="DH5(alpha)"  
/clone\_lib="605 - Endosperm cDNA library from Schmidt lab"  
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site 1: EcoRI;

```

Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 610;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGCGTGGCGACCACT 289
      |||||
Db 409 GCGGGCGTGGCGACCACT 427

RESULT 39
CB919676/c      616 bp      mRNA      linear      EST 25-APR-2003
LOCUS VVD054D06 349985 An expressed sequence tag database for abiotic
DEFINITION stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
cDNA clone VVD054D06 5, mRNA sequence.
ACCESSION CB919676
VERSION CB919676.1 GI:30134338
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 616)
AUTHORS Cushman,J.C.
TITLE An expressed sequence tag database for abiotic stressed berries of
JOURNAL Vitis vinifera var. Chardonnay
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer (backward)
Plate: 054 row: D column: 06
Seq primer: T3 20mer
High quality sequence stop: 616.
Location/Qualifiers
1. .616
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVD054D06"
/tissue_type="berries"
/dev_stage="mixed; 8, 9, 11, 13, 15, 16 weeks daf"
/clone_lib="An expressed sequence tag database for abiotic
stressed berries of Vitis vinifera var. Chardonnay"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

ORIGIN
Query Match      3.8%; Score 19; DB 14; Length 616;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCCGACGACGTGATGCG 239
      |||||
Db 24 CCGCCGACGACGTGATGCG 6

RESULT 40
AI745808      617 bp      mRNA      linear      EST 02-FEB-2000
LOCUS 605077D11.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.

```

```

AI745808      617 bp      mRNA      linear      EST 02-FEB-2000
LOCUS 605097G06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI833967
VERSION AI833967.1 GI:5456277
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 617)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605097 row: G column: 06.
Location/Qualifiers
1. .617
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: pAD-GAL4-2; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

ORIGIN
Query Match      3.8%; Score 19; DB 9; Length 617;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGCGTGGCGACCACT 289
      |||||
Db 406 GCGGGCGTGGCGACCACT 424

RESULT 41
AI833967
LOCUS 605097G06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION AI833967
VERSION AI833967.1 GI:5456277
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 617)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605097 row: G column: 06.
Location/Qualifiers
1. .617

```

```

/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5 (alpha)"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site 1: EcoRI; Site 2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 617;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGCGTGGCCGACCACT 289
414 GCCGCGTGGCCGACCACT 432

Db

RESULT 42
BM356065
LOCUS
DEFINITION
rx33b05.y1 Globodera rostochiensis J2 pCDNAII Smant v1 Globodera
rostochiensis cDNA 5' similar to TR:045080 O45080 C17H12.9 PROTEIN.
[1]; mRNA sequence.
ACCESSION
BM356065
VERSION
BM356065.1 GI:18090696
KEYWORDS
EST.
SOURCE
Globodera rostochiensis
ORGANISM
Globodera rostochiensis
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heterodoridae; Heteroderinae; Globodera.
REFERENCE
1 (bases 1 to 621)
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,
Tsagarisvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,
Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
CONTACT: McCarter, JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was contributed by Dr. Geert Smant of the Laboratory of
Nematology at Wageningen University, Wageningen,
Netherlands (geert.smant@nema.dpw.wau.nl). DNA Sequencing by:
Washington University Genome Sequencing Center

Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 420.
Location/Qualifiers
1. .621
/organism="Globodera rostochiensis"
/mol_type="mRNA"
/db_xref="taxon:31243"
/dev_stage="J2"
/lab_host="DH10B"
/clone_lib="Globodera rostochiensis J2 pCDNAII Smant v1"
/note="Vector: pCDNAII (Invitrogen); Site 1: BatXI; Site 2: EcoRI; The library was donated for sequencing by Geert Smant from Wageningen University, Laboratory of Nematology, The Netherlands."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 625;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;

Qy 178 GCCACCGTGGTGGGGCCA 196
533 GCCACCGTGGTGGGGCCA 551

Db

RESULT 43
CD428301/c
LOCUS
DEFINITION
ETH1_28_E08_g1_A002 Ethylene-treated seedlings Sorghum bicolor cDNA
clone ETH1_28_E08_A002 5', mRNA sequence.
ACCESSION
CD428301
VERSION
CD428301.1 GI:31334770
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 625)
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S.,
Klein, R.K., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R.,
Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P.,
Olaseinde, O., Eastman, A. and Pratt, L.H.
An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid
(ACC)-treated seedlings
Unpublished (2003)
Other ESTs: ETH1_28_E08.bl_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 563 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).
Location/Qualifiers
1. .625
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1_28_E08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Ethylene-treated seedlings"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from polyA+ RNA from seedlings grown
in hydroponic culture. At 8 days of age, medium was
supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic
acid (ACC) to induce endogenous ethylene (ETH) production.
Roots and shoots were harvested after 27 and 72 hr and
material from both time points was combined prior to RNA
isolation. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGG,
3-prime DraIII site is CACCAATGTG). XhoI excises the cDNA
insert."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 625;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;

Qy 178 GCCACCGTGGTGGGGCCA 196
533 GCCACCGTGGTGGGGCCA 551

Db

RESULT 43
CD428301/c
LOCUS
DEFINITION
ETH1_28_E08_g1_A002 Ethylene-treated seedlings Sorghum bicolor cDNA
clone ETH1_28_E08_A002 5', mRNA sequence.
ACCESSION
CD428301
VERSION
CD428301.1 GI:31334770
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 625)
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S.,
Klein, R.K., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R.,
Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P.,
Olaseinde, O., Eastman, A. and Pratt, L.H.
An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid
(ACC)-treated seedlings
Unpublished (2003)
Other ESTs: ETH1_28_E08.bl_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 563 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).
Location/Qualifiers
1. .625
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1_28_E08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Ethylene-treated seedlings"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from polyA+ RNA from seedlings grown
in hydroponic culture. At 8 days of age, medium was
supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic
acid (ACC) to induce endogenous ethylene (ETH) production.
Roots and shoots were harvested after 27 and 72 hr and
material from both time points was combined prior to RNA
isolation. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGG,
3-prime DraIII site is CACCAATGTG). XhoI excises the cDNA
insert."

ORIGIN
Query Match 3.8%; Score 19; DB 14; Length 625;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;

```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 327 CTCGCGCGCGAGCTCGG 345  
Db 343 CTCGCGCGCGAGCTCGG 325

RESULT 44  
CD442989/c  
LOCUS CD442989 631 bp mRNA linear EST 03-JUN-2003  
DEFINITION EL01N0420H10.b Endosperm\_4 Zea mays cDNA, mRNA sequence.  
ACCESSION CD442989  
VERSION CD442989.1 GI:31358632  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 631)  
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and Messing, J.  
TITLE Sequencing of the maize endosperm ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lai, Jinsheng  
Dr. Joachim Messing's lab  
Waksman Institute, Rutgers University  
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu  
Seq primer: T3

FEATURES  
source  
1..631  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

ORIGIN  
Query Match 3.8%; Score 19; DB 14; Length 631;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 271 GCCGGCGTGGCCGACCACT 289  
Db 452 GCCGGCGTGGCCGACCACT 434

RESULT 45  
CA400622  
LOCUS CA400622 636 bp mRNA linear EST 07-NOV-2002  
DEFINITION EL01N0407F10.g Endosperm\_4 Zea mays cDNA, mRNA sequence.  
ACCESSION CA400622  
VERSION CA400622.1 GI:24765463  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 636)  
AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and Messing, J.  
TITLE Sequencing of the maize endosperm ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lai, Jinsheng  
Dr. Joachim Messing's lab  
Waksman Institute, Rutgers University

190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jlai@waksman.rutgers.edu  
Seq primer: T7  
Location/Qualifiers  
1..636  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="W22"  
/db\_xref="taxon:4577"  
/tissue\_type="Endosperm of 7-23DAP"  
/clone\_lib="Endosperm\_4"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

ORIGIN  
Query Match 3.8%; Score 19; DB 14; Length 636;  
Best Local Similarity 100.0%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 271 GCCGGCGTGGCCGACCACT 289  
Db 438 GCCGGCGTGGCCGACCACT 456  
Search completed: June 16, 2004, 19:29:14  
Job time : 2107.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)

9488.873 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501

Sequence: 1 atgttctacagtcggcac.....cccgctcagggctcgtgc 501

Scoring table: OLIGO.NUC

Gapop 50.0 , Gapext 50.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	5.6	692	17	US-10-461-194-11
2	28	5.6	85692	17	US-10-461-194-1
3	26	5.2	88421	9	US-09-976-053-1
4	22	4.4	22	13	US-09-798-033-6
5	22	4.4	22	14	US-10-087-451-6
6	22	4.4	18876	15	US-10-329-079-42
7	22	4.4	61944	15	US-10-329-079-34
8	20	4.0	690	15	US-10-156-761-4780
9	20	4.0	1875	15	US-10-156-761-3143
10	20	4.0	1881	16	US-10-369-493-28460
11	20	4.0	1881	16	US-10-369-493-31220
12	20	4.0	64492	13	US-10-378-083-1
13	20	4.0	9025608	15	US-10-156-761-1
14	20	4.0	9025608	15	US-10-156-761-1

c 15 19 3.8 657 13 US-10-425-114-1800 Sequence 1800, Ap  
c 16 19 3.8 1172 9 US-09-924-256A-17 Sequence 17, Appl  
c 17 19 3.8 1178 9 US-09-924-256A-31 Sequence 91, Appl  
c 18 19 3.8 1212 16 US-10-369-493-40637 Sequence 40637, A  
c 19 19 3.8 1818 15 US-10-228-063-7 Sequence 7, Appl  
c 20 19 3.8 1915 13 US-10-336-753-46 Sequence 46, Appl  
c 21 19 3.8 2263 15 US-10-228-063-9 Sequence 9, Appl  
c 22 19 3.8 2263 15 US-10-272-291-1 Sequence 1, Appl  
c 23 19 3.8 2263 15 US-10-272-291-2 Sequence 2, Appl  
c 24 19 3.8 2267 10 US-09-961-077-25 Sequence 25, Appl  
c 25 19 3.8 3738 13 US-10-282-122A-13589 Sequence 13589, A  
c 26 19 3.8 4800 15 US-10-272-231-5 Sequence 5, Appl  
c 27 19 3.8 4800 17 US-10-109-048-1141 Sequence 1141, Ap  
c 28 19 3.8 11100 15 US-10-329-079-5 Sequence 5, Appl  
c 29 19 3.8 13315 15 US-10-329-079-1 Sequence 1, Appl  
c 30 19 3.8 63158 15 US-10-292-198-1 Sequence 1, Appl  
c 31 18 3.6 119 17 US-10-616-390-32 Sequence 32, Appl  
c 32 18 3.6 128 17 US-10-616-390-35 Sequence 35, Appl  
c 33 18 3.6 175 9 US-09-922-217-873 Sequence 873, App  
c 34 18 3.6 175 9 US-09-833-263-873 Sequence 873, App  
c 35 18 3.6 175 14 US-10-025-380-873 Sequence 873, App  
c 36 18 3.6 243 17 US-10-616-390-4 Sequence 4, Appl  
c 37 18 3.6 243 17 US-10-616-390-5 Sequence 5, Appl  
c 38 18 3.6 389 17 US-10-616-390-45 Sequence 45, Appl  
c 39 18 3.6 418 10 US-09-918-995-7045 Sequence 7045, Ap  
c 40 18 3.6 489 13 US-10-342-887-1619 Sequence 1619, Ap  
c 41 18 3.6 489 13 US-10-172-118-1619 Sequence 1619, Ap  
c 42 18 3.6 489 16 US-10-439-703-36 Sequence 36, Appl  
c 43 18 3.6 531 15 US-10-156-761-2447 Sequence 2447, Ap  
c 44 18 3.6 613 17 US-10-616-390-6 Sequence 6, Appl  
c 45 18 3.6 613 17 US-10-616-390-6 Sequence 6, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-461-194-11  
; Sequence 11, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-GELEDAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; PRIOR FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 692  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus

US-10-461-194-11

Query Match 5.6%; Score 28; DB 17; Length 692;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 CGCTGTTCCACAGCCTTTCGCACATCG 163  
|||  
Db 151 CGCTGTTCCACAGCCTTTCGCACATCG 178

## RESULT 2

US-10-461-194-1  
; Sequence 1, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-CELLULAMICIN PRODUCING POLYKETIDE SYNTHASES AND  
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 85692  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-1

Query Match 5.6%; Score 28; DB 17; Length 85692;  
Best Local Similarity 100.0%; Pred. No. 9.8e-05;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 CGCTGTTCCACAGCCTTTCGCACATCG 163  
|||  
Db 23581 CGCTGTTCCACAGCCTTTCGCACATCG 23608

## RESULT 3

US-09-976-059-1  
; Sequence 1, Application US/09976059  
; Patent No. US2002016474A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staiffa, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PCT  
; CURRENT APPLICATION NUMBER: US/09/976,059  
; CURRENT FILING DATE: 2001-10-15  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (6665)..(5814)  
; OTHER INFORMATION: ORF 4; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (7703)..(6693)  
; OTHER INFORMATION: ORF 5; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9464)..(8130)  
; OTHER INFORMATION: ORF 6; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9691)..(10761)  
; OTHER INFORMATION: ORF 7; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (12751)..(10829)  
; OTHER INFORMATION: ORF 8; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (13617)..(12802)  
; OTHER INFORMATION: ORF 9; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15203)..(13614)  
; OTHER INFORMATION: ORF 10; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15591)..(15863)  
; OTHER INFORMATION: ORF 11; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (15880)..(19035)  
; OTHER INFORMATION: ORF 12; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (19032)..(39713)  
; OTHER INFORMATION: ORF 13; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (39713)..(65800)  
; OTHER INFORMATION: ORF 14; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (65826)..(66530)  
; OTHER INFORMATION: ORF 15; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (66546)..(67370)  
; OTHER INFORMATION: ORF 16; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (67384)..(70059)  
; OTHER INFORMATION: ORF 17; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (70099)..(70662)  
; OTHER INFORMATION: ORF 18; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (70659)..(71906)  
; OTHER INFORMATION: ORF 19; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (73439)..(71964)  
; OTHER INFORMATION: ORF 20; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (74216)..(73563)  
; OTHER INFORMATION: ORF 21; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (75424)..(74213)

```
; OTHER INFORMATION: ORF 22; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; NAME/KEY: misc feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; positive strandedness
; NAME/KEY: misc feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1
```

```
Query Match          5.2%; Score 26; DB 9; Length 88421;
Best Local Similarity 100.0%; Pred. No. 0.001; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 20 CGACCGGGCGGCCCAAGGCGGTG 45
Db 38404 CGACCGGGCGGCCCAAGGCGGTG 38429
```

```
RESULT 4
US-09-798-033-6
; Sequence 6, Application US/09798033
; Publication No. US20020045220A1
; GENERAL INFORMATION:
; APPLICANT: Koean Biosciences, Inc.
; APPLICANT: Khosla, Chaitan
; APPLICANT: Pfeiffer, Blaine
; TITLE OF INVENTION: BIOSYNTHESIS OF POLYKETIDE SYNTHASE
; TITLE OF INVENTION: SUBSTRATES
; FILE REFERENCE: 286002021120
; CURRENT APPLICATION NUMBER: US/09/798,033
; CURRENT FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: 09/687,855
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/159,090
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/206,082
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/232,379
; PRIOR FILING DATE: 2000-09-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
```

```
; OTHER INFORMATION: Primer
US-09-798-033-6
Query Match          4.4%; Score 22; DB 13; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.73; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 ACCGAGACCTGCGGGCGATCA 460
Db 1 ACCGAGACCTGCGGGCGATCA 22

RESULT 5
US-10-087-451-6
; Sequence 6, Application US/10087451
; Publication No. US20020192767A1
; GENERAL INFORMATION:
; APPLICANT: Khosla, Chaitan
; APPLICANT: Pfeiffer, Blaine
; TITLE OF INVENTION: BIOSYNTHESIS OF POLYKETIDE SYNTHASE
; TITLE OF INVENTION: SUBSTRATES
; FILE REFERENCE: 286002021121
; CURRENT APPLICATION NUMBER: US/10/087,451
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 09/798,033
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/687,855
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/159,090
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/206,082
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 60/232,379
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/355,211
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-10-087-451-6
```

```
Query Match          4.4%; Score 22; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.73; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 ACCGAGACCTGCGGGCGATCA 460
Db 1 ACCGAGACCTGCGGGCGATCA 22

RESULT 6
US-10-329-079-42
; Sequence 42, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFEA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 18876
; TYPE: DNA
; ORGANISM: Streptomyces refuineus
```

## US-10-329-079-42

Query Match 4.4%; Score 22; DB 15; Length 18876;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CGACGGGGCGGCCCAAGGGCGT 41  
Db 13034 CGACGGGGCGGCCCAAGGGCGT 13055

## RESULT 7

US-10-329-079-34  
; Sequence 34, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
; FILE REFERENCE: 3002-11US  
; CURRENT APPLICATION NUMBER: US/10/329,079  
; CURRENT FILING DATE: 2002-12-24  
; PRIOR FILING DATE: 2001-05-30  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 34  
; LENGTH: 61944  
; TYPE: DNA  
; ORGANISM: Streptomyces refuineus  
US-10-329-079-34

Query Match 4.4%; Score 22; DB 15; Length 61944;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CGACGGGGCGGCCCAAGGGCGT 41  
Db 18094 CGACGGGGCGGCCCAAGGGCGT 18115

## RESULT 8

US-10-156-761-4780  
; Sequence 4780, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 4780  
; LENGTH: 690  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(690)  
US-10-156-761-4780

Query Match 4.0%; Score 20; DB 15; Length 690;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 GCGGCTGATCGAGCGCGAGA 256  
Db 261 GCGGCTGATCGAGCGCGAGA 280

## RESULT 9

US-10-156-761-3143  
; Sequence 3143, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3143  
; LENGTH: 1875  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1875)  
US-10-156-761-3143

Query Match 4.0%; Score 20; DB 15; Length 1875;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 ACCGGCGCGCCCAAGGGCGT 41  
Db 580 ACCGGCGCGCCCAAGGGCGT 599

## RESULT 10

US-10-369-493-28460  
; Sequence 28460, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 28460  
; LENGTH: 1881  
; TYPE: DNA  
; ORGANISM: Burkholderia fungorum  
US-10-369-493-28460

Query Match 4.0%; Score 20; DB 16; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 7 TACAGTCGGGCACGACCG 26  
|||||  
Db 709 TACAGTCGGGCACGACCG 728

## RESULT 11

US-10-369-493-31220  
; Sequence 31220, Application US/10369493  
; Publication No. US20030233675A1

## ; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 31220

; LENGTH: 1881

; TYPE: DNA

; ORGANISM: Burkholderia cepacia

US-10-369-493-31220

Query Match 4.0%; Score 20; DB 15; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TACAGTCGGGCACGACCG 26  
|||||  
Db 709 TACAGTCGGGCACGACCG 728

## RESULT 12

US-10-378-083-1

; Sequence 1, Application US/10378083

; Publication No. US20040053274A1

## ; GENERAL INFORMATION:

; APPLICANT: President of Tokyo Institute of Technology

; TITLE OF INVENTION: A gene cluster of vicenistatin biosynthesis, a viceniamine

; TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the

; TITLE OF INVENTION: polypeptide

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/378,083

; CURRENT FILING DATE: 2003-03-04

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1

; LENGTH: 64492

; TYPE: DNA

; ORGANISM: Streptomyces halstedii sp.HC-34

US-10-378-083-1

Query Match 4.0%; Score 20; DB 13; Length 64492;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ACCGGGGCGGCCCAAGGCGT 41  
|||||  
Db 42832 ACCGGGGCGGCCCAAGGCGT 42851

## RESULT 13

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

## ; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Query Match 4.0%; Score 20; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.39;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GCGGCTGATCGAGCGGAGA 256  
|||||  
Db 5836785 GCGGCTGATCGAGCGGAGA 5836804

## RESULT 14

US-10-156-761-1/c

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

## ; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Query Match 4.0%; Score 20; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.39;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ACCGGGGCGGCCCAAGGCGT 41  
|||||  
Db 3931605 ACCGGGGCGGCCCAAGGCGT 3931586



; CURRENT APPLICATION NUMBER: US/10/228,063  
; CURRENT FILING DATE: 2002-12-12  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 1818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-228-063-7

Query Match 3.8%; Score 19; DB 15; Length 1818;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1655 GCCGGCGTCCGACCACCT 1637

RESULT 20  
US-10-336-753-46/c  
; Sequence 46, Application US/10336753  
; Publication No. US20030226176A1  
; GENERAL INFORMATION:  
; APPLICANT: Guan, Hanning  
; APPLICANT: Keeling, Peter L.  
; TITLE OF INVENTION: PLANT LIKE STARCHES AND THE METHOD OF MAKING THEM IN  
; FILE REFERENCE: 2461-52  
; CURRENT APPLICATION NUMBER: US/10/336,753  
; PRIOR FILING DATE: 2003-01-06  
; PRIOR APPLICATION NUMBER: US/09/402,254  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/06660  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-03  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/042,939  
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-04-04  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 1915  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (Join(1..1815, 1819..1914))  
US-10-336-753-46

Query Match 3.8%; Score 19; DB 13; Length 1915;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1652 GCCGGCGTCCGACCACCT 1634

RESULT 21  
US-10-228-063-9/c  
; Sequence 9, Application US/10228063  
; Publication No. US20030135885A1  
; GENERAL INFORMATION:  
; APPLICANT: Lahanan, Mike  
; TITLE OF INVENTION: Self-processing Plants and Plant Parts  
; FILE REFERENCE: 109846.317  
; CURRENT APPLICATION NUMBER: US/10/228,063  
; CURRENT FILING DATE: 2002-12-12  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2223

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic  
US-10-228-063-9

Query Match 3.8%; Score 19; DB 15; Length 2223;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 2060 GCCGGCGTCCGACCACCT 2042

RESULT 22  
US-10-272-291-1/c  
; Sequence 1, Application US/10272291  
; Publication No. US20030150023A1  
; GENERAL INFORMATION:  
; APPLICANT: ExSeed Genetics  
; TITLE OF INVENTION: Starch  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/272,291  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/329,525  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2263  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Wild type sequence EX385  
US-10-272-291-1

Query Match 3.8%; Score 19; DB 15; Length 2263;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACCT 289  
|||  
Db 1845 GCCGGCGTCCGACCACCT 1827

RESULT 23  
US-10-272-291-2/c  
; Sequence 2, Application US/10272291  
; Publication No. US20030150023A1  
; GENERAL INFORMATION:  
; APPLICANT: ExSeed Genetics  
; TITLE OF INVENTION: Starch  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/272,291  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/329,525  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2263  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: mutation  
; LOCATION: (1643)  
; OTHER INFORMATION: The mutation is at position 1643 in the sequence,  
; OTHER INFORMATION: 1450 bp after the start codon.  
; FEATURE:  
; NAME/KEY: mutation  
; LOCATION: (1643)  
; OTHER INFORMATION: The change is a "G" in the wild type (EX385) to an

; OTHER INFORMATION: "A" in the mutant  
US-10-272-291-2

Query Match 3.8%; Score 19; DB 15; Length 2263;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
Db 1845 GCCGGCGTCCGACCACT 1827

RESULT 24

US-09-961-077-25/c  
; Sequence 25, Application US/09961077  
; Publication No. US20030014775A1

GENERAL INFORMATION:

APPLICANT: Zwick, Michael G.  
Edington, Brent E.  
McSwiggen, James A.  
Merlo, Patricia Ann Owens  
Guo, Lining  
Skokut, Thomas A.  
Young, Scott A.  
Polkerts, Otto  
Merlo, Donald J.

TITLE OF INVENTION: COMPOSITION AND METHODS FOR  
MODULATION OF GENE EXPRESSION  
IN PLANTS

NUMBER OF SEQUENCES: 1263

CORRESPONDENCE ADDRESS:

ADDRESSEE: LYON & LYON  
STREET: 633 West Fifth Street  
Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
Storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA: US/09/961,077

FILING DATE: 21-Sep-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/679,645

FILING DATE: July 12, 1996

APPLICATION NUMBER: 60/001,135

FILING DATE: July 13, 1995

APPLICATION NUMBER: 08/300,726

FILING DATE: September 2, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 219/247

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 2267 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-961-077-25

Query Match 3.8%; Score 19; DB 10; Length 2267;

Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
Db 1840 GCCGGCGTCCGACCACT 1822

RESULT 25

US-10-282-122A-13589  
; Sequence 13589, Application US/10282122A  
; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 13589

LENGTH: 3738

TYPE: DNA

ORGANISM: Burkholderia fungorum

US-10-282-122A-13589

Query Match 3.8%; Score 19; DB 13; Length 3738;

Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 CGGACGGCGCCGCGTCGA 491  
Db 2387 CGGACGGCGCCGCGTCGA 2405

RESULT 26

US-10-272-291-5/c

; Sequence 5, Application US/10272291

; Publication No. US20030150023A1

; GENERAL INFORMATION:

; APPLICANT: ExSeed Genetics

;; TITLE OF INVENTION: Starch  
;; FILE REFERENCE:  
;; CURRENT APPLICATION NUMBER: US/10/272,291  
;; CURRENT FILING DATE: 2002-10-17  
;; PRIOR APPLICATION NUMBER: 60/329,525  
;; PRIOR FILING DATE: 2001-10-01  
;; NUMBER OF SEQ ID NOS: 8  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 5  
;; LENGTH: 4800  
;; TYPE: DNA  
;; ORGANISM: Zea mays  
;; FEATURE:  
;; NAME/KEY: transit peptide  
;; LOCATION: (1233)..(1448)  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (1554)..(1684)  
;; OTHER INFORMATION: number 2  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (1766)..(1859)  
;; OTHER INFORMATION: number 3  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (1959)..(2054)  
;; OTHER INFORMATION: number 4  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (2145)..(2225)  
;; OTHER INFORMATION: number 5  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (2290)..(2412)  
;; OTHER INFORMATION: number 6  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (2514)..(2650)  
;; OTHER INFORMATION: number 7  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (2761)..(2857)  
;; OTHER INFORMATION: number 8  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (3102)..(3211)  
;; OTHER INFORMATION: number 9  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (3395)..(3489)  
;; OTHER INFORMATION: number 10  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (3682)..(3792)  
;; OTHER INFORMATION: number 11  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (3880)..(3976)  
;; OTHER INFORMATION: number 12  
;; FEATURE:  
;; NAME/KEY: intron  
;; LOCATION: (4106)..(4226)  
;; OTHER INFORMATION: number 13  
US-10-272-291-5

Query Match 3.8%; Score 19; DB 15; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
|||  
Db 4056 GCCGGCGTCCGACCACT 4038

RESULT 27  
US-10-109-048-1141/c  
; Sequence 1141, Application US/10109048  
; Publication No. US20040107461A1  
; GENERAL INFORMATION:  
; APPLICANT: COMMORI, PADMA  
; APPLICANT: KEELING, PETER L.  
; APPLICANT: RAMIREZ, NONA  
; APPLICANT: MCKEAN, ANGELA  
; APPLICANT: GAO, ZHONG  
; APPLICANT: GUAN, HANPING  
; TITLE OF INVENTION: GLUCAN CHAIN LENGTH DOMAINS  
; FILE REFERENCE: 2461-76  
; CURRENT APPLICATION NUMBER: US/10/109,048  
; CURRENT FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: 60/279,720  
; PRIOR FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 1154  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1141  
; LENGTH: 4800  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-109-048-1141

Query Match 3.8%; Score 19; DB 17; Length 4800;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCCGGCGTCCGACCACT 289  
|||  
Db 4056 GCCGGCGTCCGACCACT 4038

RESULT 28  
US-10-329-079-5  
; Sequence 5, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOFOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES  
; FILE REFERENCE: 3002-11US  
; CURRENT APPLICATION NUMBER: US/10/329,079  
; CURRENT FILING DATE: 2002-12-24  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 11100  
; TYPE: DNA  
; ORGANISM: Streptomyces fradiae  
US-10-329-079-5

Query Match 3.8%; Score 19; DB 15; Length 11100;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGGGCCCAAGG 38  
|||  
Db 1958 CGACCGGGGGCCCAAGG 1976

RESULT 29  
US-10-329-079-1  
; Sequence 1, Application US/10329079  
; Publication No. US20030198981A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: ZAZOFOULOS, Emmanuel  
; APPLICANT: STAFFA, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES

/ FILE REFERENCE: 3002-11US  
/ CURRENT APPLICATION NUMBER: US/10/329,079  
/ CURRENT FILING DATE: 2002-12-24  
/ NUMBER OF SEQ ID NOS: 66  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 1  
/ LENGTH: 13315  
/ TYPE: DNA  
/ ORGANISM: Streptomyces fradiae  
US-10-329-079-1

Query Match 3.8%; Score 19; DB 15; Length 13315;  
Best Local Similarity 100.0%; Pred.No. 5.6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGG 38  
Db 4173 CGACCGGGCGGCCCAAGGG 4191

RESULT 30  
US-10-292-198-1/c  
/ Sequence 1, Application US/10292198  
/ Publication No. US20030157654A1  
/ GENERAL INFORMATION:  
/ APPLICANT: SHEN, Ben  
/ APPLICANT: LIU, Wen  
/ TITLE OF INVENTION: BIOSYNTHESIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE  
/ FILE REFERENCE: 054030-0007  
/ CURRENT APPLICATION NUMBER: US/10/292,198  
/ CURRENT FILING DATE: 2003-03-14  
/ PRIOR APPLICATION NUMBER: US 10/159,257  
/ PRIOR FILING DATE: 2002-05-31  
/ PRIOR APPLICATION NUMBER: US 09/478,188  
/ PRIOR FILING DATE: 2000-01-05  
/ PRIOR APPLICATION NUMBER: US 60/115,434  
/ PRIOR FILING DATE: 1999-01-06  
/ NUMBER OF SEQ ID NOS: 146  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 1  
/ LENGTH: 63158  
/ TYPE: DNA  
/ ORGANISM: Streptomyces globisporus  
US-10-292-198-1

Query Match 3.8%; Score 19; DB 15; Length 63158;  
Best Local Similarity 100.0%; Pred.No. 3.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGG 38  
Db 42110 CGACCGGGCGGCCCAAGGG 42092

RESULT 31  
US-10-616-390-32/c  
/ Sequence 32, Application US/10616390  
/ Publication No. US20040098761A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Trick, Harold N.  
/ APPLICANT: Roe, Judith L.  
/ APPLICANT: Todd, Timothy C.  
/ APPLICANT: Herman, Michael A.  
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
/ FILE REFERENCE: KSURF-08151  
/ CURRENT APPLICATION NUMBER: US/10/616,390  
/ CURRENT FILING DATE: 2003-07-09  
/ NUMBER OF SEQ ID NOS: 45  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 32  
/ LENGTH: 119  
/ TYPE: DNA

/ ORGANISM: Heterodera glycines  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (41)..(41)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (47)..(47)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (107)..(107)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-32

Query Match 3.6%; Score 18; DB 17; Length 119;  
Best Local Similarity 100.0%; Pred.No. 54;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 101 CATGAACCGCGCGGACGG 84

RESULT 32  
US-10-616-390-35  
/ Sequence 35, Application US/10616390  
/ Publication No. US20040098761A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Trick, Harold N.  
/ APPLICANT: Roe, Judith L.  
/ APPLICANT: Todd, Timothy C.  
/ APPLICANT: Herman, Michael A.  
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
/ FILE REFERENCE: KSURF-08151  
/ CURRENT APPLICATION NUMBER: US/10/616,390  
/ CURRENT FILING DATE: 2003-07-09  
/ NUMBER OF SEQ ID NOS: 45  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 35  
/ LENGTH: 128  
/ TYPE: DNA  
/ ORGANISM: Heterodera glycines  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (24)..(24)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (52)..(52)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (84)..(84)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (90)..(90)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-35

Query Match 3.6%; Score 18; DB 17; Length 128;  
Best Local Similarity 100.0%; Pred.No. 53;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 30 CATGAACCGCGCGGACGG 47

## RESULT 33

US-09-922-217-873

; Sequence 873, Application US/09922217

; Patent No. US200200764141

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Smith, Carole Lynn

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471C13

; CURRENT APPLICATION NUMBER: US/09/922,217

; CURRENT FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 1124

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 873

; LENGTH: 175

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-922-217-873

## Query Match

3.6%; Score 18; DB 9; Length 175;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCG 314

|||||

Db 36 CCTGGTGGCGGCGCGCG 53

## RESULT 34

US-09-833-263-873

; Sequence 873, Application US/09833263

; Patent No. US20020110547A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Stolk, John A.

; APPLICANT: Meagher, Madeleine J.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND

; FILE REFERENCE: 210121.471C12

; CURRENT APPLICATION NUMBER: US/09/833,263

; CURRENT FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 1093

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 873

; LENGTH: 175

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-833-263-873

## Query Match

3.6%; Score 18; DB 9; Length 175;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCG 314

|||||

Db 36 CCTGGTGGCGGCGCGCG 53

## RESULT 35

US-10-025-380-873

; Sequence 873, Application US/10025380

; Publication No. US20020182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.

; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Smith, Carole L.

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Skeiky, Yasir A. W.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedvick Thomas S.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471C14

; CURRENT APPLICATION NUMBER: US/10/025,380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 873

; LENGTH: 175

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-025-380-873

## Query Match

3.6%; Score 18; DB 14; Length 175;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CCTGGTGGCGGCGCGCG 314

|||||

Db 36 CCTGGTGGCGGCGCGCG 53

## RESULT 36

US-10-616-390-4/c

; Sequence 4, Application US/10616390

; Publication No. US20040098761A1

; GENERAL INFORMATION:

; APPLICANT: Trick, Harold N.

; APPLICANT: Roe, Judith L.

; APPLICANT: Todd, Timothy C.

; APPLICANT: Herman, Michael A.

; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes

; FILE REFERENCE: KSURF-08151

; CURRENT APPLICATION NUMBER: US/10/616,390

; CURRENT FILING DATE: 2003-07-09

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 4

; LENGTH: 243

; TYPE: DNA

; ORGANISM: Heterodera glycines

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (145)..(145)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (151)..(151)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (183)..(183)

; OTHER INFORMATION: n is a, c, g, or t

; FEATURE:

; NAME/KEY: misc feature  
; LOCATION: (211)..(211)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-4

Query Match 3.6%; Score 18; DB 17; Length 243;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 205 CATGAACCGCGCGGACGG 188

## RESULT 37

US-10-616-390-5  
; Sequence 5, Application US/10616390  
; Publication No. US20040098761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trick, Harold N.  
; APPLICANT: Roe, Judith L.  
; APPLICANT: Todd, Timothy C.  
; APPLICANT: Herman, Michael A.  
; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
; FILE REFERENCE: KSURF-08151  
; CURRENT APPLICATION NUMBER: US/10/616,390  
; CURRENT FILING DATE: 2003-07-09  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; LENGTH: 243  
; TYPE: DNA  
; ORGANISM: Heterodera glycines  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (61)..(61)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (93)..(93)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (99)..(99)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-5

Query Match 3.6%; Score 18; DB 17; Length 243;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 39 CATGAACCGCGCGGACGG 56

## RESULT 38

US-10-616-390-45/c  
; Sequence 45, Application US/10616390  
; Publication No. US20040098761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trick, Harold N.  
; APPLICANT: Roe, Judith L.  
; APPLICANT: Todd, Timothy C.  
; APPLICANT: Herman, Michael A.  
; TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes  
; FILE REFERENCE: KSURF-08151  
; CURRENT APPLICATION NUMBER: US/10/616,390  
; CURRENT FILING DATE: 2003-07-09

; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 45  
; LENGTH: 389  
; TYPE: DNA  
; ORGANISM: Heterodera glycines  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (248)..(248)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (254)..(254)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (286)..(286)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (314)..(314)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (353)..(383)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-616-390-45

Query Match 3.6%; Score 18; DB 17; Length 389;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 CATGAACCGCGCGGACGG 479  
Db 308 CATGAACCGCGCGGACGG 291

## RESULT 39

US-09-918-995-7045  
; Sequence 7045, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7045  
; LENGTH: 418  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-7045

Query Match 3.6%; Score 18; DB 10; Length 418;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGCGCGG 367  
Db 309 GCCTGGCGCGGCGCGG 326

## RESULT 40

US-10-342-887-1619  
; Sequence 1619, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue



```
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-1619

Query Match          3.6%; Score 18; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367
Db 321 GCCTGGCGCGGGCGCGG 338

RESULT 41
US-10-172-118-1619
; Sequence 1619, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1619
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 016199
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1619

Query Match          3.6%; Score 18; DB 13; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367
Db 321 GCCTGGCGCGGGCGCGG 338

RESULT 42
US-10-439-703-36
```

```
; Sequence 36, Application US/10439703
; Publication No. US20040018527A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Jenny
; APPLICANT: O'Connell, Peter
; TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax
; FILE REFERENCE: HO-P2482US1/10205813
; CURRENT APPLICATION NUMBER: US/10/439,703
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 60/381,141
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Human
US-10-439-703-36

Query Match          3.6%; Score 18; DB 16; Length 489;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GCCTGGCGCGGGCGCGG 367
Db 321 GCCTGGCGCGGGCGCGG 338

RESULT 43
US-10-156-761-2447
; Sequence 2447, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2447
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(531)
US-10-156-761-2447

Query Match          3.6%; Score 18; DB 15; Length 531;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 CGACCTTCCTGGCGCGG 277
Db 101 CGACCTTCCTGGCGCGG 118

RESULT 44
US-10-616-390-6
; Sequence 6, Application US/10616390
; Publication No. US20040098761A1
; GENERAL INFORMATION:
```

```
/ APPLICANT: Trick, Harold N.
/ APPLICANT: Roe, Judith L.
/ APPLICANT: Todd, Timothy C.
/ APPLICANT: Herman, Michael A.
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes
/ FILE REFERENCE: KSURF-08151
/ CURRENT APPLICATION NUMBER: US/10/616,390
/ CURRENT FILING DATE: 2003-07-09
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6
/ LENGTH: 613
/ TYPE: DNA
/ ORGANISM: Heterodera glycines
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (45)..(45)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (73)..(73)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (105)..(105)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (111)..(111)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (509)..(509)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (515)..(515)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (547)..(547)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (575)..(575)
/ OTHER INFORMATION: n is a, c, g, or t
/ US-10-616-390-6
/
/ Query Match 3.6%; Score 18; DB 17; Length 613;
/ Best Local Similarity 100.0%; Pred. No. 37;
/ Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 462 CATGAACCCGCGGACGG 479
/ Db 51 CATGAACCCGCGGACGG 68
/
/ RESULT 45
/ US-10-616-390-6/c
/ Sequence 6, Application US/10616390
/ Publication No. US20040098761A1
/ GENERAL INFORMATION:
/ APPLICANT: Trick, Harold N.
/ APPLICANT: Roe, Judith L.
/ APPLICANT: Todd, Timothy C.
/ APPLICANT: Herman, Michael A.
/ TITLE OF INVENTION: Compositions and Methods for Controlling Parasitic Nematodes
/ FILE REFERENCE: KSURF-08151
/ CURRENT APPLICATION NUMBER: US/10/616,390
/ CURRENT FILING DATE: 2003-07-09
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6
```

```
/ LENGTH: 613
/ TYPE: DNA
/ ORGANISM: Heterodera glycines
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (45)..(45)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (73)..(73)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (105)..(105)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (111)..(111)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (509)..(509)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (515)..(515)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (547)..(547)
/ OTHER INFORMATION: n is a, c, g, or t
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (575)..(575)
/ OTHER INFORMATION: n is a, c, g, or t
/ US-10-616-390-6
/
/ Query Match 3.6%; Score 18; DB 17; Length 613;
/ Best Local Similarity 100.0%; Pred. No. 37;
/ Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 462 CATGAACCCGCGGACGG 479
/ Db 569 CATGAACCCGCGGACGG 552
/
/ Search completed: June 16, 2004, 20:03:04
/ Job time : 265 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 Seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325  
Perfect score: 501  
Sequence: 1 atgtttacagtcggcac.....ccgcgctcgggctgtgc 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 521

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	4.0	5275	1	US-08-485-588-1
2	20	4.0	5275	1	US-08-484-565-1
3	20	4.0	5275	2	US-08-480-751-1
4	20	4.0	5275	2	US-08-943-986-1
5	20	4.0	5275	3	US-08-353-784-1
6	20	4.0	5275	3	US-08-484-719B-1
7	20	4.0	5275	4	US-08-484-159-1
8	19	3.8	1172	3	US-08-861-774E-17
9	19	3.8	1178	4	US-08-861-774E-91
10	19	3.8	1818	4	US-09-731-166-3
11	19	3.8	2267	4	US-08-679-645-25
12	19	3.8	4800	3	US-08-941-443A-4
13	18	3.6	495	4	US-09-621-976-1819
14	18	3.6	497	4	US-09-621-976-797
15	18	3.6	498	2	US-08-722-349-2
16	18	3.6	498	3	US-09-204-328-2
17	18	3.6	789	4	US-09-252-991A-6590
18	18	3.6	942	4	US-09-252-991A-6619
19	18	3.6	954	4	US-09-252-991A-8902
20	18	3.6	2469	4	US-09-252-991A-6679
21	18	3.6	3111	4	US-09-252-991A-8790
22	18	3.6	77536	4	US-09-410-551B-1
23	18	3.6	4403765	3	US-09-103-840A-2
24	18	3.6	4411529	3	US-09-103-840A-1
25	17	3.4	333	4	US-09-060-756-318
26	17	3.4	333	4	US-09-670-314-318
27	17	3.4	456	4	US-09-252-991A-9308

ALIGNMENTS

RESULT 1

US-08-485-588-1  
; Sequence 1, Application US/08485588  
; Patent No. 5688938  
; GENERAL INFORMATION:  
; APPLICANT: Edward M. Brown  
; APPLICANT: Steven C. Hebert  
; APPLICANT: Forrest H. Fuller  
; APPLICANT: James E. Garrett, Jr.  
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
; MOLECULES  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: First Interstate World Center  
; STREET: Suite 4700  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,588  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 9  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451

Sequence 680, App  
Sequence 680, App  
Sequence 9256, Ap  
Sequence 4264, Ap  
Sequence 2, Appli  
Sequence 15003, A  
Sequence 6751, Ap  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 6874, Ap  
Sequence 4787, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 9291, Ap  
Sequence 9280, Ap  
Sequence 1, Appli  
Sequence 2, Appli

```
/
/ FILING DATE: 23 August, 1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Heber, Sheldon O.
/ REGISTRATION NUMBER: 38,179
/ REFERENCE/DOCKET NUMBER: 213/005
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5275 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna to mRNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 515..3769
/ OTHER INFORMATION:
/ US-08-485-588-1
/
/ Query Match 4.0%; Score 20; DB 1; Length 5275;
/ Best Local Similarity 100.0%; Pred. No. 2.5;
/ Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 362 GCGCGGTCTCGCGCGGG 381
/ Db 223 GCGCGGTCTCGCGCGGG 242
/
/ RESULT 2
/ US-08-484-565-1
/ Sequence 1, Application US/08484565
/ Patent No. 5763569
/ GENERAL INFORMATION:
/ APPLICANT: Edward M. Brown
/ APPLICANT: Steven C. Hebert
/ APPLICANT: James E. Garrett, Jr.
/ TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
/ MOLECULES
/ NUMBER OF SEQUENCES: 20
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: First Interstate World Center
/ STREET: Suite 4700
/ STREET: 633 West Fifth Street
/ CITY: Los Angeles
/ STATE: California
/ COUNTRY: USA
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: FASTSEQ
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/484,565
/ FILING DATE: 7 June, 1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ PRIOR APPLICATION DATA: including application
/ PRIOR APPLICATION DATA: described below: 9
/ APPLICATION NUMBER: 08/353,784
/ FILING DATE: 9 December, 1994
/ APPLICATION NUMBER: PCT/US/94/12117
/ FILING DATE: 21 October, 1994
/ APPLICATION NUMBER: U.S. 08/292,827
/ FILING DATE: 23 August, 1994
/ APPLICATION NUMBER: U.S. 08/141,248
/ FILING DATE: 22 October, 1993
/ APPLICATION NUMBER: U.S. 08/009,389
/ FILING DATE: 23 February, 1993
/
```

```
/
/ APPLICATION NUMBER: U.S. 08/017,127
/ FILING DATE: 12 February, 1993
/ APPLICATION NUMBER: U.S. 07/934,161
/ FILING DATE: 21 August, 1992
/ APPLICATION NUMBER: U.S. 07/834,044
/ FILING DATE: 11 February, 1992
/ APPLICATION NUMBER: U.S. 07/749,451
/ FILING DATE: 23 August, 1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Heber, Sheldon O.
/ REGISTRATION NUMBER: 38,179
/ REFERENCE/DOCKET NUMBER: 213/006
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5275 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cdna to mRNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 515..3769
/ OTHER INFORMATION:
/ US-08-484-565-1
/
/ Query Match 4.0%; Score 20; DB 1; Length 5275;
/ Best Local Similarity 100.0%; Pred. No. 2.5;
/ Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 362 GCGCGGTCTCGCGCGGG 381
/ Db 223 GCGCGGTCTCGCGCGGG 242
/
/ RESULT 3
/ US-08-480-751-1
/ Sequence 1, Application US/08480751
/ Patent No. 5858684
/ GENERAL INFORMATION:
/ APPLICANT: Edward F. Nemeth
/ APPLICANT: Edward M. Brown
/ APPLICANT: Steven C. Hebert
/ APPLICANT: Forrest H. Fuller
/ APPLICANT: James E. Garrett, Jr.
/ TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
/ MOLECULES
/ NUMBER OF SEQUENCES: 20
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: First Interstate World Center
/ STREET: Suite 4700
/ STREET: 633 West Fifth Street
/ CITY: Los Angeles
/ STATE: California
/ COUNTRY: USA
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: FASTSEQ
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/480,751
/ FILING DATE: 7 June, 1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ PRIOR APPLICATION DATA: including application
/ PRIOR APPLICATION DATA: described below: 9
/ APPLICATION NUMBER: 08/353,784
/
```

; FILING DATE: 9 December, 1994  
; APPLICATION NUMBER: PCT/US/94/12117  
; FILING DATE: 21 October, 1994  
; APPLICATION NUMBER: U.S. 08/292,827  
; FILING DATE: 23 August, 1994  
; APPLICATION NUMBER: U.S. 08/141,248  
; FILING DATE: 22 October, 1993  
; APPLICATION NUMBER: U.S. 08/009,389  
; FILING DATE: 23 February, 1993  
; APPLICATION NUMBER: U.S. 08/017,127  
; FILING DATE: 12 February, 1993  
; APPLICATION NUMBER: U.S. 07/934,161  
; FILING DATE: 21 August, 1992  
; APPLICATION NUMBER: U.S. 07/834,044  
; FILING DATE: 11 February, 1992  
; APPLICATION NUMBER: U.S. 07/749,451  
; FILING DATE: 23 August, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heber, Sheldon O.  
; REGISTRATION NUMBER: 38,179  
; REFERENCE/DOCKET NUMBER: 213/004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5275 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 515..3769  
; OTHER INFORMATION:  
; US-08-480-751-1

Query Match 4.0%; Score 20; DB 2; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGGGTCTCGGCGCGGG 381  
DB 223 GCGGGTCTCGGCGCGGG 242

RESULT 4  
US-08-943-986-1  
; Sequence 1, Application US/08943986  
; Patent No. 5962314  
; GENERAL INFORMATION:  
; APPLICANT: Edward M. Brown  
; APPLICANT: Steven C. Hebert  
; APPLICANT: James E. Garrett, Jr.  
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: First Interstate World Center  
; STREET: Suite 4700  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FASTSEQ  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/943,986  
; FILING DATE: 03-OCT-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/484,565  
; FILING DATE: 7-June-1995  
; APPLICATION NUMBER: 08/353,784  
; FILING DATE: 9 December, 1994  
; APPLICATION NUMBER: PCT/US/94/12117  
; FILING DATE: 21 October, 1994  
; APPLICATION NUMBER: U.S. 08/292,827  
; FILING DATE: 23 August, 1994  
; APPLICATION NUMBER: U.S. 08/141,248  
; FILING DATE: 22 October, 1993  
; APPLICATION NUMBER: U.S. 08/009,389  
; FILING DATE: 23 February, 1993  
; APPLICATION NUMBER: U.S. 08/017,127  
; FILING DATE: 12 February, 1993  
; APPLICATION NUMBER: U.S. 07/934,161  
; FILING DATE: 21 August, 1992  
; APPLICATION NUMBER: U.S. 07/834,044  
; FILING DATE: 11 February, 1992  
; APPLICATION NUMBER: U.S. 07/749,451  
; FILING DATE: 23 August, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heber, Sheldon O.  
; REGISTRATION NUMBER: 38,179  
; REFERENCE/DOCKET NUMBER: 213/006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5275 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 515..3769  
; OTHER INFORMATION:  
; US-08-943-986-1  
Query Match 4.0%; Score 20; DB 2; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 362 GCGGGTCTCGGCGCGGG 381  
DB 223 GCGGGTCTCGGCGCGGG 242  
RESULT 5  
US-08-353-784-1  
; Sequence 1, Application US/08353784  
; Patent No. 6011068  
; GENERAL INFORMATION:  
; APPLICANT: Edward F. Nemeth, Edward M.  
; APPLICANT: Brown, Steven C. Hebert,  
; APPLICANT: Bradford C. Van Wagenen, Manuel  
; APPLICANT: F. Balandrin, Forrest H. Fuller,  
; APPLICANT: Eric G. DelMar, and Scott T. Moe  
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: First Interstate World Center  
; STREET: Suite 4700  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles

STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,784  
FILING DATE: 9 December, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application described below: 8  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 209/069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 515..3769  
OTHER INFORMATION:  
US-08-353-784-1  
Query Match 4.0%; Score 20; DB 3; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 362 GCGGGTCTCGGCGCGGG 381  
Db 223 GCGGGTCTCGGCGCGGG 242  
RESULT 6  
US-08-484-719B-1  
Sequence 1, Application US/08484719B  
Patent No. 6031003  
GENERAL INFORMATION:  
APPLICANT: Edward F. Nemeth, Edward M.  
APPLICANT: Brown, Steven C. Hebert,  
APPLICANT: Bradford C. Van Wagenen,  
APPLICANT: Manuel F. Balandrin,  
APPLICANT: Forrest H. Fuller, Eric G.  
APPLICANT: Delmar, Scott T. Moe  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

TITLE OF INVENTION: MOLECULES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS Word  
SOFTWARE: FASTSEQ for Windows Version 3.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,719B  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Douglas C. Murdock  
REGISTRATION NUMBER: 37,549  
REFERENCE/DOCKET NUMBER: 213/007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 515..3769  
OTHER INFORMATION:  
US-08-484-719B-1  
Query Match 4.0%; Score 20; DB 3; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 362 GCGGGTCTCGGCGCGGG 381  
Db 223 GCGGGTCTCGGCGCGGG 242  
RESULT 7  
US-08-484-159-1  
Sequence 1, Application US/08484159  
Patent No. 6313146  
GENERAL INFORMATION:

APPLICANT: Bradford C. Van Wagenen  
APPLICANT: Manuel F. Balandrin  
APPLICANT: Eric G. Del Mar  
APPLICANT: Edward F. Nemeth  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
MOLECULES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,159  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 9  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 214/101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 515..3769  
OTHER INFORMATION:  
US-08-484-159-1

Query Match 4.0%; Score 20; DB 4; Length 5275;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCTCGGCGCGGG 381  
|||||

Db 223 GCGCGGTCTCTCGGCGCGGG 242

## RESULT 8

US-08-861-774E-17/c  
; Sequence 17, Application US/08861774E  
; Patent No. 6297007  
; GENERAL INFORMATION:  
; APPLICANT: Waters, Barbara  
; APPLICANT: Miao, Vivian  
; APPLICANT: Ho, Yap  
; APPLICANT: Tong, Seow  
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR  
; BIOACTIVE MOLECULES  
; FILE REFERENCE: 9993-006  
; CURRENT APPLICATION NUMBER: US/08/861,774E  
; CURRENT FILING DATE: 1997-05-22  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 1172  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps32  
US-08-861-774E-17

Query Match 3.8%; Score 19; DB 3; Length 1172;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCTCGGCGAC 23  
|||||

Db 1171 TCTACACGTCTCGGCGAC 1153

## RESULT 9

US-08-861-774E-91/c  
; Sequence 91, Application US/08861774E  
; Patent No. 6297007  
; GENERAL INFORMATION:  
; APPLICANT: Waters, Barbara  
; APPLICANT: Miao, Vivian  
; APPLICANT: Ho, Yap  
; APPLICANT: Tong, Seow  
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR  
; BIOACTIVE MOLECULES  
; FILE REFERENCE: 9993-006  
; CURRENT APPLICATION NUMBER: US/08/861,774E  
; CURRENT FILING DATE: 1997-05-22  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 91  
; LENGTH: 1178  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps25  
US-08-861-774E-91

Query Match 3.8%; Score 19; DB 3; Length 1178;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TCTACACGTCTCGGCGAC 23  
|||||

Db 1177 TCTACACGTCTCGGCGAC 1159

## RESULT 10

US-09-731-166-3/c  
; Sequence 3, Application US/09731166  
; Patent No. 6639126

```
/
/ GENERAL INFORMATION:
/ APPLICANT: Sewalt, Vincent J. H.
/ APPLICANT: Singletary, George W.
/ TITLE OF INVENTION: Production of Modified Polysaccharides
/ FILE REFERENCE: 35718/206348
/ CURRENT APPLICATION NUMBER: US/09/731,166
/ CURRENT FILING DATE: 2000-12-06
/ PRIOR APPLICATION NUMBER: 60/169,993
/ PRIOR FILING DATE: 1999-12-06
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 1818
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (0)...(0)
/ OTHER INFORMATION: UDP-glucose starch glycosyl transferase -- Genbank
/ OTHER INFORMATION: Accession No. 6639126 22509
/ NAME/KEY: CDS
/ LOCATION: (1)...(1818)
/
US-09-731-166-3

Query Match          3.8%; Score 19; DB 4; Length 1818;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGCGTGGCGACCACT 289
DB 1652 GCGGGCGTGGCGACCACT 1634

RESULT 11
US-08-679-645-25/c
Sequence 25, Application US/08679645
Patent No. 6350934
GENERAL INFORMATION:
APPLICANT: Zwick, Michael G.
APPLICANT: Edington, Brent E.
APPLICANT: McSwigen, James A.
APPLICANT: Merlo, Patricia Ann Owens
APPLICANT: Guo, Lining
APPLICANT: Skokut, Thomas A.
APPLICANT: Young, Scott A.
APPLICANT: Folkerts, Otto
APPLICANT: Merlo, Donald J.
TITLE OF INVENTION: COMPOSITION AND METHODS FOR
TITLE OF INVENTION: MODULATION OF GENE EXPRESSION
TITLE OF INVENTION: IN PLANTS
NUMBER OF SEQUENCES: 1263
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,645
FILING DATE: July 12, 1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/001,135
FILING DATE: July 13, 1995
APPLICATION NUMBER: 08/300,726
```

```
/
/ FILING DATE: September 2, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Warburg, Richard J.
/ REGISTRATION NUMBER: 32,327
/ REFERENCE/DOCKET NUMBER: 219/247
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 25:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2267 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-08-679-645-25

Query Match          3.8%; Score 19; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGGGCGTGGCGACCACT 289
DB 1840 GCGGGCGTGGCGACCACT 1822

RESULT 12
US-08-941-445A-4/c
Sequence 4, Application US/08941445A
Patent No. 6107060
GENERAL INFORMATION:
APPLICANT: Keeling, Peter
APPLICANT: Guan, Haining
TITLE OF INVENTION: Starch Encapsulation
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/941,445A
FILING DATE: 30-SEP-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,855
FILING DATE: 30-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Winner, Ellen P
REGISTRATION NUMBER: 28,547
REFERENCE/DOCKET NUMBER: 89-97
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4800 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
```



LOCATION: join(1449..1553, 1685..1765, 1860..1958, 2055  
 LOCATION: ..2144, 2226..2289, 2413..2513, 2651..2760, 2858  
 LOCATION: ..3101, 3212..3394, 3490..3681, 3793..3879, 3977  
 LOCATION: ..4105, 4227..4343)  
 US-08-941-445A-4

Query Match 3.6%; Score 19; DB 3; Length 4800;  
 Best Local Similarity 100.0%; Pred. No. 7.3;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 GCCGGCGTCCGACACCT 289  
 Db 4056 GCCGGCGTCCGACACCT 4038

RESULT 13

US-09-621-976-1819  
 ; Sequence 1819, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 1819  
 ; LENGTH: 495  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 21..335  
 ; NAME/KEY: sig.peptide  
 ; LOCATION: 21..191  
 ; OTHER INFORMATION: Von Heijne matrix  
 ; OTHER INFORMATION: score 4.4000009536743  
 ; OTHER INFORMATION: seq STLCTAPLSTCG/EP  
 US-09-621-976-1819

Query Match 3.6%; Score 18; DB 4; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 350 GCCTGGCGCGGGCGCGG 367  
 Db 338 GCCTGGCGCGGGCGCGG 355

RESULT 14

US-09-621-976-797  
 ; Sequence 797, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 797  
 ; LENGTH: 497  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 182..496

US-09-621-976-797

Query Match 3.6%; Score 18; DB 4; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 350 GCCTGGCGCGGGCGCGG 367  
 Db 330 GCCTGGCGCGGGCGCGG 347

RESULT 15

US-08-722-349-2  
 ; Sequence 2, Application US/08722349  
 ; Patent No. 5955299  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.  
 ; APPLICANT: Bandman, Olga B.  
 ; APPLICANT: Zweiger, Gary B.  
 ; TITLE OF INVENTION: NOVEL SNRNP SM PROTEINS  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: U.S.  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/722,349  
 ; FILING DATE: Filed Herewith  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0132 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-855-0555  
 ; TELEFAX: 415-845-4166  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 498 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cdna  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY:  
 ; CLONE: Consensus  
 US-08-722-349-2

Query Match 3.6%; Score 18; DB 2; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 350 GCCTGGCGCGGGCGCGG 367  
 Db 339 GCCTGGCGCGGGCGCGG 356

RESULT 16

US-09-204-328-2  
 ; Sequence 2, Application US/09204328  
 ; Patent No. 6090564  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hillman, Jennifer L.

APPLICANT: Bandman, Olga  
APPLICANT: Zweiger, Gary B.  
TITLE OF INVENTION: NOVEL SNRP SM PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/204,328  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,349  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0132 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 498 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE: Consensus  
US-09-204-328-2

Query Match 3.6%; Score 18; DB 3; Length 498;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 350 GCCTGGCCGGCGGCGG 367  
Db 339 GCCTGGCCGGCGGCGG 356

RESULT 17  
US-09-252-991A-6590  
; Sequence 6590, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6590  
; LENGTH: 789  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6590  
Query Match 3.6%; Score 18; DB 4; Length 789;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 266 TCCTGGCCGGCGTCCCGA 283  
Db 159 TCCTGGCCGGCGTCCCGA 176

RESULT 18  
US-09-252-991A-6619  
; Sequence 6619, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6619  
; LENGTH: 942  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6619

Query Match 3.6%; Score 18; DB 4; Length 942;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 TCCTGGCCGGCGTCCCGA 283  
Db 776 TCCTGGCCGGCGTCCCGA 793

RESULT 19  
US-09-252-991A-8902  
; Sequence 8902, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8902  
; LENGTH: 954  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8902

Query Match 3.6%; Score 18; DB 4; Length 954;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 GCGGATCGGCGTCCCGG 359  
Db 735 GCGGATCGGCGTCCCGG 752

RESULT 20  
US-09-252-991A-6679/c  
; Sequence 6679, Application US/09252991A

; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6679  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6679

Query Match 3.6%; Score 18; DB 4; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 TCCTGGCGCGGTGCCGA 283  
|||  
DB 2045 TCCTGGCGCGGTGCCGA 2028

RESULT 21  
US-09-252-991A-8790  
; Sequence 8790, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8790  
; LENGTH: 3111  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8790

Query Match 3.6%; Score 18; DB 4; Length 3111;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 GCGGATCGGCTGGCCGG 359  
|||  
DB 2146 GCGGATCGGCTGGCCGG 2163

RESULT 22  
US-09-410-551B-1/c  
; Sequence 1, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00

; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 77536  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (52275)...(71465)  
US-09-410-551B-1

Query Match 3.6%; Score 18; DB 4; Length 77536;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 CCGCTGTTCACAGCCT 152  
|||  
DB 45947 CCGCTGTTCACAGCCT 45930

RESULT 23  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 3.6%; Score 18; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 CCGCGACGGCGCGCG 487  
|||  
DB 719415 CCGCGACGGCGCGCG 719398

RESULT 24  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 3.6%; Score 18; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 CGCGGACGGCGCGCG 487  
DB 717982 CGCGGACGGCGCGCG 717965

RESULT 25  
US-09-060-756-318/c  
; Sequence 318, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 318  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-060-756-318

Query Match 3.4%; Score 17; DB 3; Length 333;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACGGGTGCT 125  
DB 71 GACCAGGACGGGTGCT 55

RESULT 26  
US-09-670-314-318/c  
; Sequence 318, Application US/09670314  
; Patent No. 6492506  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/670,314  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 09/060,756  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 318  
; LENGTH: 333  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-670-314-318

Query Match 3.4%; Score 17; DB 4; Length 333;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACGGGTGCT 125  
DB 71 GACCAGGACGGGTGCT 55

RESULT 27  
US-09-252-991A-9308  
; Sequence 9308, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9308  
; LENGTH: 456  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9308

Query Match 3.4%; Score 17; DB 4; Length 456;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCTGCGCGGGGC 363  
DB 138 TCGGCTGCGCGGGGC 154

RESULT 28  
US-09-060-756-680/c  
; Sequence 680, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 680  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (various positions within the sequence)  
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"  
US-09-060-756-680

Query Match 3.4%; Score 17; DB 3; Length 507;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACCGGTGCT 125  
|||  
DB 93 GACCAGGACCGGTGCT 77

## RESULT 29

US-09-670-314-680/c  
; Sequence 680, Application US/09670314  
; Patent No. 6492506  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/670,314  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 09/060,756  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 680  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (various positions within the sequence)  
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"  
US-09-670-314-680

Query Match 3.4%; Score 17; DB 4; Length 507;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 GACCAGGACCGGTGCT 125  
|||  
DB 93 GACCAGGACCGGTGCT 77

## RESULT 30

US-09-252-991A-9256/c  
; Sequence 9256, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9256  
; LENGTH: 768  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9256

Query Match 3.4%; Score 17; DB 4; Length 768;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCTGGCCGGGGGC 363  
|||  
DB 749 TCGGCTGGCCGGGGGC 733

## RESULT 31

US-09-252-991A-4264  
; Sequence 4264, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4264  
; LENGTH: 909  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4264

Query Match 3.4%; Score 17; DB 4; Length 909;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 GCGGATCGGCTGGCG 358  
|||  
DB 513 GCGGATCGGCTGGCG 529

## RESULT 32

PCT-US95-04801-2  
; Sequence 2, Application PC/TUS9504801  
; GENERAL INFORMATION:  
; APPLICANT: Martin, Juan F.  
; APPLICANT: Coque, Juan R.  
; APPLICANT: Enguita, Francisco J.  
; APPLICANT: Fuente, Juan L.  
; APPLICANT: Llarena, Francisco J.  
; APPLICANT: Liras, Paloma  
; TITLE OF INVENTION: DNA ENCODING CEPHAMYCIN BIOSYNTHESIS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John W. Wallen III  
; STREET: P.O. Box 2000  
; CITY: Rahway  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07065  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04801  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wallen III, John W.  
; REGISTRATION NUMBER: 35,403  
; REFERENCE/DOCKET NUMBER: 19179  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-3905  
; TELEFAX: (908) 594-4720  
; INFORMATION FOR SEQ ID NO: 2;

SEQUENCE CHARACTERISTICS:  
LENGTH: 972 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-04801-2

Query Match 3.4%; Score 17; DB 5; Length 972;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 GCGCGGGGCGCGGTCC 370  
DB 954 GCGCGGGGCGCGGTCC 970

RESULT 33  
US-09-252-991A-15003  
Sequence 15003, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15003  
LENGTH: 999  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15003

Query Match 3.4%; Score 17; DB 4; Length 999;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 CCGCGGAGCGGTGATG 237  
DB 281 CCGCGGAGCGGTGATG 297

RESULT 34  
US-09-252-991A-6751  
Sequence 6751, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6751  
LENGTH: 1062  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6751

Query Match 3.4%; Score 17; DB 4; Length 1062;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 ACCACTGGTGGGGCC 309  
DB 503 ACCACTGGTGGGGCC 519

RESULT 35  
US-08-529-600D-1  
Sequence 1, Application US/08529600D  
Patent No. 5861285  
GENERAL INFORMATION:  
APPLICANT: Tadashi MATSUNAGA  
TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND  
TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,600D  
FILING DATE: 18-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-248700  
FILING DATE: 16-SEP-1994  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1302  
US-08-529-600D-1

Query Match 3.4%; Score 17; DB 2; Length 1302;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCCTGGCGCGGTG 279  
DB 971 CCTTCCTGGCGCGGTG 987

RESULT 36  
US-08-973-275-5  
Sequence 5, Application US/08973275B  
Patent No. 5958706  
GENERAL INFORMATION:  
APPLICANT: MATSUNAGA, Tadashi  
APPLICANT: KAMIYA, Shinji  
APPLICANT: NAKABA, Kenryo  
TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS  
TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: MATSUNAGA  
CURRENT APPLICATION NUMBER: US/08/973,275B  
CURRENT FILING DATE: 1998-02-09  
EARLIER APPLICATION NUMBER: PCT/JP97/01043  
EARLIER FILING DATE: 1997-03-27  
EARLIER APPLICATION NUMBER: JP 8-97536  
EARLIER FILING DATE: 1996-03-28

Thu Jun 17 09:02:22 2004

```

; EARLIER APPLICATION NUMBER: JP 8-146833
; EARLIER FILING DATE: 1996-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 5
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Magnetospirillum AMB-1
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(1302)
; OTHER INFORMATION: Genomic DNA sequence encoding mag A gene.
US-08-973-275-5

Query Match          3.4%; Score 17; DB 2; Length 1302;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCTGCGCGCGGTG 279
Db 971 CCTTCTGCGCGCGGTG 987

RESULT 37
US-09-122-632-1
; Sequence 1, Application US/09122632
; Patent No. 6033878
; GENERAL INFORMATION:
; APPLICANT: Tadaashi MATSUNAGA
; TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND
; TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: PILLSBURY MADISON & SUTRO, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/122,632
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/529,600
; FILING DATE: 18-SEP-1995
; APPLICATION NUMBER: JP 6-248700
; FILING DATE: 16-SEP-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1302
US-09-122-632-1

Query Match          3.4%; Score 17; DB 3; Length 1302;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 CCTTCTGCGCGCGGTG 279
Db 971 CCTTCTGCGCGCGGTG 987
```

```

RESULT 38
US-09-252-991A-6874/c
; Sequence 6874, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6874
; LENGTH: 1530
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6874

Query Match          3.4%; Score 17; DB 4; Length 1530;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 ACCACCTGGTGGCGGCC 309
Db 716 ACCACCTGGTGGCGGCC 700

RESULT 39
US-09-489-039A-4787/c
; Sequence 4787, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4787
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4787

Query Match          3.4%; Score 17; DB 4; Length 1587;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 GCCGCCCGCGCAGCGCG 323
Db 273 GCCGCCCGCGCAGCGCG 257

RESULT 40
US-07-820-011A-1
; Sequence 1, Application US/07820011A
; Patent No. 5336615
; GENERAL INFORMATION:
; APPLICANT: Bell, Leonard
; APPLICANT: Madri, Joseph A.
; APPLICANT: Warren, Stephen L.
; APPLICANT: Luthringer, Daniel J.
; TITLE OF INVENTION: Genetically Engineered
; TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced
; TITLE OF INVENTION: Migration
```

Thu Jun 17 09:02:22 2004

TITLE OF INVENTION: and Plasminogen Activator Activity

NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Maurice M. Klee  
 STREET: 1951 Burr Street  
 CITY: Fairfield  
 STATE: Connecticut  
 COUNTRY: USA  
 ZIP: 06430  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 Kb storage  
 COMPUTER: IBM PC XT  
 OPERATING SYSTEM: PC-DOS/MS-DOS 2.10  
 SOFTWARE: Displaywrite 3  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/820,011A  
 FILING DATE: 19920106  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Klee, Maurice M.  
 REGISTRATION NUMBER: 30,399  
 REFERENCE/DOCKET NUMBER: 1B-101  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (203) 255 1400  
 TELEFAX: (203) 254 1101  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1602 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: Double  
 TOPOLOGY: Linear  
 MOLECULE TYPE: cdna to mRNA  
 HYPOTHETICAL: No  
 ANTI-SENSE: No  
 ORIGINAL SOURCE:  
 ORGANISM: Gallus, gallus  
 PUBLICATION INFORMATION:  
 AUTHORS: Takeya, Tatsuo  
 TITLE: Structure and Sequence of the  
 TITLE: Cellular Gene Homologous to the RSV src  
 TITLE: Gene and the Mechanism for Generating the  
 TITLE: Transforming Virus  
 JOURNAL: Cell  
 VOLUME: 32  
 PAGES: 881-890  
 DATE: March, 1983  
 PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 1; Length 1602;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 GCACCGAGACCTGCGGG 453  
 DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 41  
 PCT-US93-00445-1  
 Sequence 1, Application PC/TUS9300445  
 GENERAL INFORMATION:  
 APPLICANT: Bell, Leonard  
 APPLICANT: Madri, Joseph A.  
 APPLICANT: Warren, Stephen L.  
 APPLICANT: Luthringer, Daniel J.  
 TITLE OF INVENTION: Genetically Engineered  
 TITLE OF INVENTION: Endothelial Cells  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Maurice M. Klee  
 STREET: 1951 Burr Street  
 CITY: Fairfield

STATE: Connecticut  
 COUNTRY: USA  
 ZIP: 06430  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch, 760 Kb storage  
 COMPUTER: DELL 486/50  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Displaywrite 3  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/00445  
 FILING DATE: 19930105  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/820,011  
 FILING DATE: 06-JAN-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Klee, Maurice M.  
 REGISTRATION NUMBER: 30,399  
 REFERENCE/DOCKET NUMBER: ALX-101PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (203) 255 1400  
 TELEFAX: (203) 254 1101  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1602 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: Double  
 TOPOLOGY: Linear  
 MOLECULE TYPE: cdna to mRNA  
 HYPOTHETICAL: No  
 ANTI-SENSE: No  
 ORIGINAL SOURCE:  
 ORGANISM: Gallus, gallus  
 PUBLICATION INFORMATION:  
 AUTHORS: Takeya, Tatsuo  
 TITLE: Structure and Sequence of the  
 TITLE: Cellular Gene Homologous to the RSV src  
 TITLE: Gene and the Mechanism for Generating the  
 TITLE: Transforming Virus  
 JOURNAL: Cell  
 VOLUME: 32  
 PAGES: 881-890  
 DATE: March, 1983  
 PCT-US93-00445-1

Query Match 3.4%; Score 17; DB 5; Length 1602;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 GCACCGAGACCTGCGGG 453  
 DB 1149 GCACCGAGACCTGCGGG 1165

RESULT 42  
 US-09-252-991A-9291  
 Sequence 9291, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 9291  
 LENGTH: 1629



ATTORNEY/AGENT INFORMATION:  
NAME: Wallen III, John W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-04801-1

Query Match 3.4%; Score 17; DB 5; Length 1700;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 GGCCTGGCGCGGTC 370  
DB 54 GGCCTGGCGGTC 70

RESULT 45  
US-09-470-881-2  
Sequence 2, Application US/09470881  
Patent No. 6685938  
GENERAL INFORMATION:  
APPLICANT: CHERESH, David A.  
APPLICANT: ELICEIRI, Brian  
TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL FOR MODULATION OF  
TITLE OF INVENTION: ANGIOGENESIS AND VASCULAR PERMEABILITY USING SRC OR  
TITLE OF INVENTION: YES TYROSINE KINASES  
FILE REFERENCE: TSRI 651.2  
CURRENT APPLICATION NUMBER: US/09/470,881  
CURRENT FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US99/11780  
PRIOR FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: 60/087,220  
PRIOR FILING DATE: 1998-05-29  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1759  
TYPE: DNA  
ORGANISM: Chicken  
FEATURE:  
NAME/KEY: gene  
LOCATION: (1)..(1759)  
OTHER INFORMATION: chicken c-SRC cDNA  
NAME/KEY: CDS  
LOCATION: (112)..(1710)  
US-09-470-881-2

Query Match 3.4%; Score 17; DB 4; Length 1759;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 GCACCGAGACCTGGCGG 453  
DB 1260 GCACCGAGACCTGGCGG 1276

Search completed: June 16, 2004, 19:32:35  
Job time : 70.3333 secs

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9291

Query Match 3.4%; Score 17; DB 4; Length 1629;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCTGGCGGGGC 363  
DB 944 TCGGCTGGCGGGGC 960

RESULT 43  
US-09-252-991A-9280/c  
Sequence 9280, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 9280  
LENGTH: 1641  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9280

Query Match 3.4%; Score 17; DB 4; Length 1641;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 TCGGCTGGCGGGGC 363  
DB 852 TCGGCTGGCGGGGC 836

RESULT 44  
PCT-US95-04801-1  
Sequence 1, Application PC/TUS9504801  
GENERAL INFORMATION:  
APPLICANT: Martin, Juan F.  
APPLICANT: Coque, Juan R.  
APPLICANT: Enguita, Francisco J.  
APPLICANT: Fuente, Juan L.  
APPLICANT: Lliarena, Francisco J.  
APPLICANT: Liras, Paloma  
TITLE OF INVENTION: DNA ENCODING CEPHAMYCIN BIOSYNTHESIS  
TITLE OF INVENTION: LATE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John W. Wallen III  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04801  
FILING DATE:  
CLASSIFICATION:

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:02:16 ; Search time 2074.33 Seconds  
(without alignments)  
10468.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_1825\_2325

Perfect score: 501  
Sequence: 1 atgttctacacgtcgggcac.....cccggtcgagggtcgtgc 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 15  
Total number of hits satisfying chosen parameters: 8137

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pt.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_vl.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
4	28	5.6	69644	1	AY179507	AY179507 Streptomy
5	26	5.2	82746	1	AF453501	AF453501 Actinosyn
6	26	5.2	88421	6	AX417445	AX417445 Sequence
7	24	4.8	32870	1	AF007101	AF007101 Streptomy
8	24	4.8	347750	1	AP002998	AP002998 Mesorhizo
9	23	4.6	43254	1	MLCB1779	MLCB1779 Mycobacteri
10	23	4.6	312050	1	MLEPRTN3	MLEPRTN3 Mycobacteri
11	22	4.4	7887	6	BD262946	BD262946 DNA sequ
12	22	4.4	9858	1	AF321122	AF321122 Streptomy
13	22	4.4	9858	6	AX803758	AX803758 Sequence
14	22	4.4	18876	6	AF324838	AF324838 Streptomy
15	22	4.4	34869	1	BD262936	BD262936 DNA sequ
16	22	4.4	37856	6	AX024212	AX024212 Sequence
17	22	4.4	37856	1	AF322256	AF322256 Streptomy
18	22	4.4	39428	1	AX803750	AX803750 Sequence
19	22	4.4	61944	6	EX664737	EX664737 Homo sapi
20	21	4.2	144560	2	EX664737	EX664737 Homo sapi
21	21	4.2	189050	1	AL646066	AL646066 Ralstonia
22	21	4.2	219447	9	HUMFLNG6PD	HUMFLNG6PD Ralstonia
23	21	4.2	349305	1	EX640433	EX640433 Bordetell
24	20	4.0	5249	4	S67307	S67307 Ca (2+) - sens
25	20	4.0	5275	6	AR012622	AR012622 Sequence
26	20	4.0	5275	6	AR028465	AR028465 Sequence
27	20	4.0	5275	6	AR078217	AR078217 Sequence
28	20	4.0	5275	6	AR177781	AR177781 Sequence
29	20	4.0	5275	6	I75051	I75051 Sequence 1
30	20	4.0	11842	1	AE005053	AE005053 Halobacte
31	20	4.0	23451	1	AB070952	AB070952 Streptomy
32	20	4.0	28890	1	AF512431	AF512431 Saccharot
33	20	4.0	77292	8	EX842596	EX842596 Neurospor
34	20	4.0	102842	8	AF497474	AF497474 Aegilops
35	20	4.0	203050	1	AL646071	AL646071 Ralstonia
36	20	4.0	208524	1	AE016925	AE016925 Chromobac
37	20	4.0	299800	1	AP005040	AP005040 Streptomy
38	20	4.0	300900	1	AP005939	AP005939 Bradyrhiz
39	20	4.0	302300	1	AP005034	AP005034 Streptomy
40	19	3.8	558	6	AX756416	AX756416 Sequence
41	19	3.8	579	8	AF267643	AF267643 Zea mays
42	19	3.8	716	1	AF481794	AF481794 Streptomy
43	19	3.8	1150	8	AF079237	AF079237 Arundinel
44	19	3.8	1172	6	AR170948	AR170948 Sequence
45	19	3.8	1178	6	AR170983	AR170983 Sequence

ALIGNMENTS

RESULT 1	AMM223012	53784 bp	DNA	linear	BCT 09-FEB-1998
LOCUS	AMM223012				
DEFINITION	Amycolatopsis mediterranei genes encoding rifamycin polyketide synthases, ORFs 1 to 5.				
ACCESSION	AJ223012				
VERSION	AJ223012.1	GI:2764760			
KEYWORDS	ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.				
SOURCE	Amycolatopsis mediterranei				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.				
REFERENCE	1				
AUTHORS	Schupp, T., Toupet, C., Engel, N. and Goff, S.				
TITLE	Cloning and sequence analysis of the putative rifamycin polyketide				

synthase gene cluster from *Amycolatopsis mediterranei*  
 Unpublished  
 2 (bases 1 to 53784)  
 Schupp, T.  
 Direct Submission  
 Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
 AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
 SWITZERLAND

FEATURES  
 source  
 1. 53784  
 /organism="Amycolatopsis mediterranei"  
 /mol\_type="genomic DNA"  
 /strain="IBG A3136"  
 /db\_xref="taxon:33910"  
 1336..15543  
 /function="CoA ligase, acylcarrier protein,  
 b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"  
 Unpublished  
 /notes="ORF1  
 PKS modules 1-3"  
 /codon\_start=1  
 /transl\_table=11  
 /product="trifamycin polyketide synthase, type 1"  
 /protein\_id="CA11035.1"  
 /db\_xref="GI:2764761"  
 /db\_xref="GOA:O54666"  
 /db\_xref="SPTREMBL:O54666"  
 /translations="MRTDLIPLRLHVALLENATRFAGKPAFPADHRTVTYGDLEARTRR  
 LAGHLAGLVGRHGRVNAICLGNRVSTVSYFAILRAGVGVPLNPGSATALEHPLTD  
 SGATVYVDAQAARLEALRVELLVGTDDVPSGAHSYDELASEPAEAADLEUDE  
 PAMWVFTGTGRPKGVVSTORCLMSVASCYVPFGLSDQDRVLPFLPHLSHIA  
 CVLSATVGVASVRIADGSSADDVMRLIEASSTFLAGVPTTYHHLVRAAKRQGSFAPS  
 LRTCLAGACVLAGLRSSEFETFGVLIDAYSTETCGAITMNPDPGARVQSGSCGLAV  
 PDGVVVPDDVTGLDPVAGBEGVSPGNMVLGYHNSPEATAAARDGWDFRTGDLAR  
 RDDAGVYTI CGRIKELIIRGGANTHGCEVAELRTVDGVADAAGVGVDPDITLGEVPA  
 YVTPGCPDPAALIEKCRQLSNKYVPDRLLEVAHPTASCKIRGGLLTDRPAOLR  
 YAATEHEEGSRHADESVAAALRLKSLGDERACECELLEDLVRTQAADVLGQVPDGEA  
 FRDIFGTSLAIVELRNRLTSTGLWLPASAVDFDHTPAAALRAVRLGILVQVAPAE  
 VRAADPCGETAIVGMAICRLPGVVASPEDLWRLVAERVDVASEFFPGRGMDDSLIDPD  
 REAGTYSVYQGGFLHDAGEFDAGFVGSIPREAVAMPDQORLLLETSWALEANAGVDP  
 IALKGTDGVFSGLMGQYSGSVAPELEGFTVTGVASSVASGRVSVTLGLEGPATV  
 DTACSSISVAMHLLAQAALROGECSMALLAGGVYTMATPGSFVEPSRORALAPDCRKAF  
 AAAAGDTGWSEGVGVVLRISVARERGHRIILAVLRGSAVNDGASNGLTAPNGLSQ  
 RIVTRRALAAGAPSDVVDVYAEHGTGTTGLDPIEAQALLATYGOERQKPLMLGKNG  
 IGHQAAGAVAGVYIKMQALRHETLPTLTHVDKPLEVDWSAGAIELLTFAEWAPKRN  
 RPRAGVSYSGVCTNAHLILEEAPAEPEVAAPELPVTVLVVARSATSESLSGAERLA  
 SILEGDLTFEVAQALSRRAVLDERAVVWAGSREEAVTGLRLANTAGSGTGGKVVWV  
 PPGOGTQAGMGRLELLAESVPFAERIAECALAPAWIDMSI.VDVLRGEGDLGRVDLQ  
 PACFAVMVGLAAWVESVGVRPDAVGHSGQGEIAAACVSGALSLEDAAKVALLRSQAIA  
 AELSGRGMSALVGDGVDVSRLLVDGVEVAANGPSSVITAGDAHALDAILLELSGBG  
 IRVRVAVDYASHTRHVEDIRTLAETLAGISAQAPAVPYFTVTSWSRDAGVLDGG  
 WYVRLNRQVFGAAATALLPQCHTVFVEVSAHPTVQPLSEGTDAIGTLRRDEGLG  
 RLLASWCELFLVGGIDVDMTAMVPAAGWDI.LPTVA.FEHRHWLBPASPAGAPDLGCT  
 VVTFPGSDRLTAVQWSSRRQAQPWADGLVFNAAIVEAAI.RLGLDIAFTVGVGLVVDAP  
 WLPFRGSRFQVLIVGEPEGQRRLPIEVFSREADEPFRHAHGTLIAAAA.VPEPAAA  
 GDATDVTAGLRDADRQVHLLDAAVTVGTTDDLLPSSVWMTVSLASGATAVTFVTE  
 TATGLRLDPAGQVPLTVESVRGPTFVAEQCTDILFRFDWPEIPLPTAEATDLVTE  
 ATSAEATLSALQAMLDPAETRLAVVTGDCTEPGAAA.IWGLVRSQAQSEHPRI.VIADL  
 DDPAVLPAVVASGEPQVRVNGVAVPEL.TRTVTPROQADPLDPEGTVLITGGCTGLGA  
 LTLRALHVAHGRHLVLVSRRGAPLEQELTLGASVA.IAACVDVARQALEAVLRAL  
 PAEHLPTAVITHAGVDAANASDELALRQNRNSLDLFAVSI.AMGVYVDSVTEHGHGDA  
 SSAAGVLPQPGQAAYANASDELALRQNRNSLDLFAVSI.AMGVYVDSVTEHGHGDA  
 LRRNRIGMSGLPADGNEMLDAAIATGCTTLVAAKFDVAALRAXAGAGVPPPLGLRI  
 APLPERAAKTASLTERLAGLAETEQAAALLDLVRHAAAEVLHSGHSAESVHSGRTFKL  
 AGFDSLITAVELRNELAAATGLTSPAMI.FDYKPPALADHLRAKLFGSAA.NRPABI.GYI  
 AABEPTIAIVMAKCPFGGVHSPEDLWRLVADGADAVTDFPADRGMDITLHYEDPOH  
 EETTVYRHGAFILDDAAGDAFAFGISPEALAMDEPQKLLLETSWELFERAAIIDPTTL  
 ACQDITGVAGVNSHDSNMRHRAAGVEGFRLTGSGSALVTSGRVAFHGFGEPAVTVTDI  
 ACSLSLVAHMAVQALROGECSMALGAGVMTVETVFEFSRQRLADPGRCRKAFL  
 GADGTGWEGVGLLIVLRSRQGRGHQVIAVVRGSAVNSDASNGLTAPNPGSPQOR  
 TKALAAAGLSTSDVDAVEAHTGTLTGDPIEA EALLATYQNRLEPLTARQWVNSNGHL  
 ITHQAAGVAGVYIKVMAMRHGVLPRTLHVRDPSSTVDMSAGAVELLTARQWVNSNGHL  
 REAGVSYIGTGNHAFVLEEVAAPI.TTTPQPEPAEFLVPLVARSATAGLGRQAGRLA  
 AFLGDRDTVVRDPAAYALATRAQLDHRAVVLVLSADRAQLCADLAAFGSGVTVTGPDV



[illegible]



```

/product="putative secreted protein"
/db_xref="AA006910.1"
/db_xref="GI:28192595"
/translation="MDKRMGHRHRTTQPRRTTLATRAVLAAGVLTPTIASAGSAHAA
TPQAAICTDRDELAKDSEINSALEGAATATSLHRTTNTCTLDADRHFDAS
TVKVTVLISLNDQAOKNALTQEEKHATAMITSDNDATTALMKQLGADKINFLQ
AAGMTITLDSSEHGLQITANDDEKLLQLVHTNPVLSDDSEKYLILKTAEVIPSQ
RWGTGAPADQAQVKNGLERATNGVHSLGATGCGDHDYTTITVLSQDNATWDDG
IANIEGIARAHENLNAPVSSAQ"
complement(1945..2736)
/notes="similar to S. coelicolor SCF1.09 (SC0267); ORF18"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/db_xref="AA006911.1"
/db_xref="GI:28192596"
/translation="MTIRVPAGSLDVRVGMGTPLVFAHYMGSAADTWNGVIGHL
PQCATVDFQRMGSRPALPGYHLIDQLADDLVRIEVRVSGPFLVGHSMGKAGQ
LVAAARPAGLVGLVAPAPPPAPPPMRYEYRQGLSHAYDSAEVTHALDHVITATPL
PEAVRATAERDSLAAAGAEARQEWPLRGVIARDITDAARRIEVTVVLAANDVVEPPhi
LRHLPHIPHTLITVPDAGHLLPAEAPGAVARALGDFLTATQS"
2830..3774
/notes="Arac family; similar to S. coelicolor SC5A7.14
(SC6644) and SCF1.08c (SC0266); ORF19"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="AA006912.1"
/db_xref="GI:28192597"
/translation="MENGSAVQRQLRVLPAVGAYGVEVLDFAALRSMQTRRTQPO
PLGFHVALVSGTSGHEADPHNYRLGEGGAVIRPQWHRWSDIDACDGPFLILFRP
RPSFGTASEATAPACWHLDRQLSLALIAHLGREHSTAVHTPLASPVLSHLAA
LILRALGTPSPGSPSPRPTEVFRAYRAAVERFTDWHVADYARALGYDVRTIT
RTTAAATGAKFLDQRIILLEAKLLAHTDLPVSGCARLGRDVCNFTFFPRQAG
LPPAAARAAVSTQAQCG"
complement(3824..4585)
/notes="TetR family; similar to S. coelicolor 2SCD46.08
(SC4194); ORF20"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="AA006913.1"
/db_xref="GI:28192598"
/translation="MLSVNTVTPSHLNAPTLRTGKYPGAEIYQGFSAFVADVKG
PGLERKQTHAAVDAASLISLELYGPNQVSVQAABAAEVSKRTLPAYPTKEDLV
VHLADHETESAVRAAPPTAPLTALREHLSGLRDEPITGLNDQPEVRLHRMI
VLDASLVAMAGFKTGABRALTEALRETPELTARLAAQVIAVHWAQAQDNAERL
AYGPADRYAGADVADMEHAFALLENGRLHLPQ"
4569..6065
/genes="gdmI"
/notes="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11
/product="GdmI"
/db_xref="AA006914.1"
/db_xref="GI:28192599"
/translation="MLTSTTVVAGATGCLMAYELALAGVETLVLEKLPQRIQQ
VKGQITQRTAELESGLLEPMLRAIARDVPVGSFGALPVLDCARAHYLVACDGGH
IPQWIEEVEERATAGARVLRGTAVGVADDGVVTVADGLRCAHYLVACDGGH
STVKILGLPFGAGTHPAVLADIRLSAVSLVPRQMGSLMTFRHARGVWSMLVPL
GDVRYFTGHADQADTADPTVTHEEIAAALQAVYPTTLGADVNSRFDATROL
EYHRTGRLPAGDAHILPGLAQGLNGLQDALNGLKWLAAVLQDAPRNGDLSVHA
EHPVAAQVHTTSQRVLAINSPEDVAALRDIPTDLRLPDTNRHLAGLSLSR
YDLPDGHPTGRIIPDADLVETGTITSLTSLFGSSHAVLLDLAGVADPLPLPPRVDL
VRATCADDMAALIRPDGYVCWATDTSAAACDITLLAALTGLARVP"
complement(6107..6577)
/genes="gdmX"
/notes="similar to JdxX and MmyY"
/codon_start=1

```

```

/transl_table=11
/product="GdmX"
/protein_id="AA006915.1"
/db_xref="GI:28192600"
/translation="MTATADYDSLARRIALRLEKDALRALMLRGRWALDRKQWTFIE
CWAEADLEFGWEKIHGKEAVRAKVEVESYQSMQHHILNMHWVDGEDRATGIGYM
WFAVATPGSTSSPYSMGPGPYDWEPRRPGDWLLVRQLGVMTDGEDTLRAFA"
7208..27736
/genes="gdmA1"
/notes="polyketide synthase; modules 0-3"
/codon_start=1
/transl_table=11
/product="GdmA1"
/protein_id="AA006916.1"
/db_xref="GI:28192601"
/translation="MLRTDLIRPVPPELLRANADRFQKPCASDGHRTVSHAELRRTR
RIAGHLAQLRHPGORAMICLGNRVMVESYFGLFANGVAVPNRSTDAELSYLLA
DSGARLVLDVAHADQFGLRQFPELRVVVSGDGLPKGFIATFGLPTEPQLAARD
DGLDEITWMLTSGTGLPKGVLSQTNCLSLAACVYVPTGLTAEDRVLPLPFLH
SUSHVCLLAATAVAGASTRIVDGVSTADVLDALEERSFIAGVPTLYHHLIABAER
DSTATPELIALVGGAVATADLVRSFESAFGLVPLVDAYSTETCGATANVWPGPRVEG
SCGLPVGLTIRLVDDTGDVPAGEGEFVWGNMAGIYHGMATASALRDXNR
TGBLGRDAGCTVTTRIKELVIRAGENIHPEGEVAVLRTFVGVADAAYVGRPHAVL
GEVPAFVPGDGEDPSALLATCRERLSYKVPPEIYEIARVPTASGKITRHLVLE
LPLRAAGDQYDLSLLRDWVQPSALPDAPACTGTWALADADALGLAVRAAGVDA
RVGPEVPEFVADSVAGLVAGSVADLAGDGAAPDVVVVTPPVAGIDETGAPDEAGV
TVGERADRLAARLGAWLADRLAGTTFVVAATGAVATGAEDAPEPLSALWGVRSLL
QAAVPRGLTIVDVLGDAGDAGDAGEDREARLRAVGGHQAIRGGLVLPKL
OAYSPFAEPGAPPTLDAGLVITGTGTRTALARHLVAYGARNLLSANGLPPE
TTSFPAEPGAPPTLDAGLVITGTGTRTALARHLVAYGARNLLSANGLPPE
AANTLRELDAGNOVMAVCDPADRTALDSVLDAQAPVTAHVHIEEPRESLDTG
LRAMHLERTRGAAAFVYVTSAGVLSGPRGPDAADAOCEALRRRRLGLG
LAWLWGLPGEHGTAPVAGAVPLPEALALFDALTAGCQGLVLLRPSGTGPGPEVP
AVLRHIVDAPSPASDEPAVAFRRRLAASESGRQRMALALVREHAATLGLASD
PVBAGQAFSAFGFTSLTAVARNLNAATCARLAATVVDHPTTAPGARHLRVRITGR
RGVQAPVARGVSDPEVAIVAMGCHLPEGEVATPEDLWRLVADGRDALAGPDEKMDL
AGLSDSDPARGVSVREGGFLTDAGGFDAFFGISPREALAMPQOGLLETWETF
ENAGIDPSLHGTDVGFSGVMVDYGDAGTAEGLEHGLVGSAGSVVSGRVAYAL
GLTPAVTDTACSSSLVALHVAQVTRTCECSLALAGVAVNSRPTSFIFRSRGL
APDRCKSFAGADGTNWSGVGLVLLERLSDARRNGHEVLAVVRGTAIVNQDGSNGL
TAPNGSQERVIRQALANAGTVADVDVAEHAGTGTSLGDPIDEAQALLATYQGRPED
OPLWGLSKNIGHAQAAGAGVIMVQAMHGVLPKTLHADEPTSKDWTSGAVSL
LSEARPECTHPRAGISSPGVSTNAHVLEQALEAAAPETASDAGAPAGVATG
GVVFWLSAKTIPALRAQARLUSHLESSESDANVDVCSLATTALEHRAVILATD
AEGMAVALAEGRPOPLNVITGTDGKTVPFIPEGCAQWGMGQALLNTSPVFAA
RLRECDALAPYTDWLSLIDVITGTPDPSLERVVDVQPAIFAVVSLAALLWQSGIHP
DAVHGSQGEITAAACVAGHLTNAAKITVLRSTQIIAHLHAGHGMMSLATPDIIDL
TNWKGKWIARHNSPNATVIAGDTDALHQLHLYTDQGTARAIIPVDYASHTGVDTI
KNQLODVLGDTLEPGTIPMLSTVTGOWIEPNTVGDSSVYVNRNQTVQFERTIHTLAD
QGYRTYIEISPHVLTALQETLEANDPTNTITVGLRRDDDTPTRLTNLAHLTN
GTPVNVPLFTGTQTPRIPLPTYPFQHHLWLPNTSAGDVSAGVLCQGTGHPLAGAVV
SVPTGVLTLTQSLATHPAWADHVSQTVLLPSTAMELAIRAGDSDTPTLELV
IGQPMTLPEDGALHVQVLVGVDEGRGVRVYSPDAAQOEWEHSHAGSTLQTDQDGS
AEGMENGMPWPPGVEPIALDDDFDOLAQGYEYGFAPGLKRWKRDGDEVFABAA
LPEEDTVAGRFGIHPALLDAALHNSFCVPPAPGQTLIPFVWNGVRLILACATVRV
BARATGDTFTLSLDSGPSVASVDSVLRALISPEOLAAASGGAGRADALFTLDWT
EHPALGTEVSWAILGDAHTDVADHVALIAGEORPGAVADTAAMAAGDTGLPARAR
DLARALDLVQRVGRPELDVRLVLLTRGAVSHDTEAVTDPAADAINGLVRSQASE
HPRALVDVTDVDSREALPEVAAGEPQVALRGLLWVRLVRSQGLLAPVAHEWYL
DVSEKSLNVLVRPPEATPLATQVRIEYRAAQONFRDVLVGLGVSAGGQGGGGE
GAGVYTEVPGVGLAVGDVWGLPFRGFIADARTVPIPEGWVSATVAAQVFA
VLTALYGLRDTGTPGGETVLAAGVGMAAVOLARHFGATVATVAHPSKHVLTA
LGVPEGHLLSGDLGFAAPALVDVLSLTCVYDASLGLLGTGGRFVEMKNDIRD
PASVAALHPGVGYQAPFDLGGDAGPVRIRELLABELVELPEAGRIEPLFIHWUVOQPT
AFRMSQGRHTGKIVITLPRALDPDGTVLITGGTGTGLTARWVTHHGAQRLLIS
ROGPAPAGDTLTBLTBLGATVITACDTRDQDLAALADIPAAHPLTAVIHTAGA
LDDGYLTALTPRDLTPRKVDVAVTHLDLTDQDLAAFYSSAAGTILNAGQANY
AAANAFIDAPQRHARHARPATSLAWGJWSDTSLTSTMDATDVRTRAGVGLMDNA
LSGQLAGLSPAGQREFLLNVRAEAVALHAGPEALEPTVAFKEMGFDSLTAVELRN
RINAATGLRLPATLLFDTLPTALLTFLTEHGGGPAPAAAAPVTVRAAADPIAVVA
MSCRLPGGVTDPTDGLWNLLLEERDGIADPRDRGWDLEALFDADPRSGTSVYLRGGF

```

gene

CDS

CDS

CDS

CDS

gene

CDS

gene

CDS







/product="RNA polymerase sigma factor"  
/protein\_id="AM54109.1"  
/db\_xref="GI:21449377"  
/translation="WMIGYSDGRRRRPVRWGWTEATFTWETDEDIKLAATLRDAEA  
ANADSTRAYLKAIGRVLRLVVSIAKRYTGKMAILLDPPPAAGKRRPVELR  
RVVRQQRKHLRLANLLVVSIAKRYTGKMAILLDPPPAAGKRRPVELR  
FKFYSYAWIRIQATRALADQSRTIRLPHVWVVTKLARLERDLNREATAAE  
VAVELDVPEKVELROHARTVSLDQITGAEDTSIGDLIAADENAVPLEASALL  
SADLRVLETLTPREAGVVRMYGFPDGVHLTLDDEVGRVYGVTRIRQIESRTWXL  
QHPARSGLGYLD"  
complement (6501..9518)  
/gene="asm30"  
complement (6501..9518)  
/gene="asm30"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450"  
/protein\_id="AM54108.1"  
/db\_xref="GI:21449376"  
/translation="WVATGTRIPKPLVGNLLDLVLSLTDLDVDFLDRCHREHG  
IVALTAGQKQVPSHSELVARNCSPPSKAVHPALEQVRDPAGDGLFTARGDEPNW  
GKARLLMPAGFPAMRDHPFAMLDIAEQMLVWRPFGEDHRIDVADMTLTLDTIA  
LCAGFARNSFYDRAPHFVDAMRVSLVEAGRAELPGQVFLVGRNQYRDDIATM  
NRADGVAAALPAGERDDLLERMLTCADPTGERSANRVYQLATFLTAGHET  
TSLGSAVRLIAHPVELKAKADVGLGDRVPAPEDLARLDYLGVLRLTLLHP  
TAPALAPPEALGSHAGCEPVLMPLTLHRDPVWRDPDVPDPERFAPERWDE  
IPACWMPFGHGRACIGRFPALQENTLVLLVLRQFDLALADPDHRLTIKQTLTKP  
DSLIVRARPRPGATATVETVPHQVPAIHRHGTPHLYVFGSGSGGLGRTIAG  
DGAARWATSVLDDRALPASGPVIVSSSYNGAPPDAAHFVKWLQDGPDLISG  
VDYLVLCGNLMSATYQRYPTLIDEMAAGARLRERGAATDARADFFGDWERYEP  
LWPLSACEVEIGEPRFVRVSDAAGDLGLASAVLENRELVRGPDAGSKRHLE  
LRPDGTSYGTGVLSPQNPDLVRAVARLCTRAERVVTVSSAPDGLVPGVGRAL  
RVDELLTRCDLSAPAGGVVARLAERCPCPPERAELAAITGTLLELLERPGCAVD  
LALALELLPRTLYISIAESAQRAEVALTVSTGTVSGYLSERVGRVAVGAS  
PPESFPADNPVLTAGTGIAPFRGLRAALGEGPCPALLLFGCGPELDDL  
YAEFFALGWLVDRAYSRHPDGEVHRVQRLRRDRVRELVDAGARVYLCGDATR  
VGPAVEVLGRIGPGAGWLDALRAGRYATDVF"  
9664..10245  
/gene="asm29"  
9664..10245  
/gene="asm29"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AM54107.1"  
/db\_xref="GI:21449375"  
/translation="MSAVRRQLVAAALVWRKDCIAAATTAICAEADMPHGAFY  
CFRSQELYTALLTTGLTSLDAWPAISPDLPVNVKTMALAEFTWPAQADPERLV  
LGLVTLAQRPELRLPAWQHREYVAKAVAHLDLAEAGPDYSVDRLALBLVVS  
LDGVASSWFSDHDELAKASLDQFAVMEALTR"  
10393..10857  
/gene="asm28"  
10393..10857  
/gene="asm28"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AM54106.1"  
/db\_xref="GI:21449374"  
/translation="MTDITTRHTFVQNAALAVGAVFLVGIAGFVPGLTNTYDGLF  
AGHSDAALLGVFKYSALNVLHIAFGAAGIAAGLSKAPRGAFLIGGAIYAVLWL  
YGVVDEHSTANFVPLNADNMLHLGLAVGMTIGLGAALGRNAHYRAARSDD"  
complement (10912..11937)  
/gene="asm27"  
complement (10912..11937)  
/gene="asm27"  
/codon\_start=1  
/transl\_table=11  
/product="glucose 1-dehydrogenase"  
/protein\_id="AM54105.1"  
/db\_xref="GI:21449373"  
/translations="MPRLAEQVVVVVVVGVASSGGTGRVTLRFAAAGARVVCARNTRAL  
DGLVEVRGAGGRAVATDAEAAVRAADVAVRFGVDTWVNAAGTGVGRVED

TPAGEFDEVMVNLGHVHGAKAALPALRRAGGVLIIGVASVGLRSAPLQAPYAASK

Query Match 5.2%; Score 26; DB 1; Length 82746;  
Best Local Similarity 100.0%; Pred. No. 0.088;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 ATCGACGCTACGGCAGCAGCAGCAGAC 446  
|||  
DB 36138 ATCGACGCTACGGCAGCAGCAGCAGAC 36113  
|||

RESULT 6  
AX417445 88421 bp DNA linear PAT 18-JUN-2002  
LOCUS Sequence 1 from Patent WO0231155.  
DEFINITION AX417445  
ACCESSION AX417445  
VERSION AX417445.1 GI:21522731  
KEYWORDS  
SOURCE Actinoplanes sp.  
ORGANISM Actinoplanes sp.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Actinoplanes.

REFERENCE 1  
AUTHORS Zazopoulos E., Farnet, C.M. and Staffa, A.  
TITLE Gene cluster for ramoplanin biosynthesis  
JOURNAL Patent: WO 0231155-A 1 18-APR-2002;  
ECOPTA BIOSCIENCES INC (CA)  
FEATURES  
1..88421  
Location/Qualifiers  
source  
/organism="Actinoplanes sp."  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1871"  
misc\_feature  
/note="ORF 1; positive strandedness"  
2077..3078  
misc\_feature  
/note="ORF 2; positive strandedness"  
3118..4032  
misc\_feature  
/note="ORF 3; positive strandedness"  
4038..5048  
misc\_feature  
/note="ORF 4; negative strandedness"  
complement (5814..6665)  
misc\_feature  
/note="ORF 5; negative strandedness"  
complement (6693..7703)  
misc\_feature  
/note="ORF 6; negative strandedness"  
complement (8130..9464)  
misc\_feature  
/note="ORF 7; positive strandedness"  
9691..10761  
misc\_feature  
/note="ORF 8; negative strandedness"  
complement (10829..12751)  
misc\_feature  
/note="ORF 9; negative strandedness"  
complement (12802..13617)  
misc\_feature  
/note="ORF 10; negative strandedness"  
complement (13614..15203)  
misc\_feature  
/note="ORF 11; positive strandedness"  
15591..15863  
misc\_feature  
/note="ORF 12; positive strandedness"  
15880..19035  
misc\_feature  
/note="ORF 13; positive strandedness"  
19032..39713  
misc\_feature  
/note="ORF 14; positive strandedness"  
39713..65800  
misc\_feature  
/note="ORF 15; positive strandedness"  
65826..66530  
misc\_feature  
/note="ORF 16; positive strandedness"  
66546..67370  
misc\_feature  
/note="ORF 17; positive strandedness"  
67384..70059  
misc\_feature  
/note="ORF 18; positive strandedness"  
70099..70662  
misc\_feature  
/note="ORF 19; positive strandedness"  
70659..71906  
misc\_feature  
/note="ORF 20; negative strandedness"  
complement (71964..73439)  
misc\_feature  
complement (73563..74216)



AFDPDPFGISPREALAMDPOORLLLETTWEVLERAGIDPTILGRSRTGVTFTGIMYODY  
VTRLRDPFALGNGSAAVASGRVSVTFEGEPVSLDTACSSLSVALHVAVRP  
ALRRLARVAGGVTVMSVPHYPVEFSRQGLAFDARCKSPDASADGTIVLSEGVGLLIY  
ERUSDALRGHEVLAVVRGAVNQDASNGLTAFPTGFSQQRVIRDAITDARLTVADV  
LLEAHGTGRLGDEIEAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGVKTVM  
AMRGHPTLVSEPTPNVDSAGVRLTWARDFSTDQRRRAAVSFGISGTNA  
HVLEQAPAEPEVDVPTPAGVPWVLGAHTGPALSALAEALAAHEVDEAPLGDA  
HLENTPVLAALRAVVRVEEVAALRAVPRTRRGRPLAVMTGQSGSRFGMGRO  
LVEHFPFAQLDEVALATPGIREVMEFDPOQARTLQRTDHAQIALPAETALRYLNE  
SWGLRDMVCHSVGHITAHVSGTLLTDDAHLVITRGILMONLPFGAMLAIVADP  
HTQLPHLDNHHTDITISAAINGPHATVLSGRTTLTHIAQTNTKTNWLNASHAFSPFL  
MLOLPQFTTLTTHPHPHPTPLISMLTATPTDPTHTTQWITPAVRYTDTLHLH  
HGHITYLEGPTTLTALATPLTPTLTHLIPTRNHNVEVSTNEALGRVFSVGHV  
DMRALPTGRTSLPTVFORRDFWLHDAAGGAEGVAGGLGTTHDPLGLGAVTADT  
GELLSGRSTSHPLTHTVNGTVIPTGALIDLAHAAHDTHTVDELVIHTPL  
ALHTPSPQSTVGAETDTHRPALASRDITGTWTRHTGTLSNQTHPAAELTNSPT  
DAHQIDLTAYQLADGLHYGPAQGLHTLYQDNTLTFABIELEPAAAGLTGHILHP  
ALIDAAHPAAVAARDEESGILPFSWAGTVHSTGATHRVRLDTDGTGARRLAT  
DITGQVPIIDSLTTRPLHDPQHLAAQEPNTHLYTLGWSPTPTPTTGTGHERYTVPT  
AGDIIADTHRIITAOGLTRIQOHLAADTTPLVVEATHDDLACARVWGLRTAQTEHP  
GRILIDTNHPTSRDALASVAGKEQARIDGALTLPRLTSLDVTASTRSTEDL  
SMGFLVITGTTGGLSGLIARHIVTEHGLDLVLVSRQGPADGATLTLQLOQHGARI  
RTISCDUTDHTLTLIDELGTLGVVHTAGALADTTIDHNLNPDITLTPTPKANPAW  
WLHELTQDHLALVFFSVVAGVGNAGQANYAANSFLDALTTHRRROGLPSTLAW  
GMWRGEGTAHLTQADHDMTRAGIHLGTLDAEGTTLFDTLDRGLGAVAPVKHLPT  
LNRAADTVAVLRGLVRAVPAARQATVPAESWADRIAALMPEERDRAVLDRVQOQAR  
VLGSHDADRDPGAKFDGTGFSLTAVEPRNVTTATGLRLPATLIFDHPPTNALADH  
LLAQGLPLGDSGSSPVRDLRERFERQLFDRPDDVEGTALAVRLSLAHLDAFAT  
APAAATASQVLDLTEDASOSELISFIDKEFGASN"  
14964..21266  
/note="orf3"  
/codon\_start=1  
/transl\_table=11  
/product="PKS module 3"  
/protein\_id="AAC38063.1"  
/db\_xref="GI:2624950"

CDS

21363..27176  
/note="orf4"  
/codon\_start=1  
/transl\_table=11  
/product="PKS module 4"  
/protein\_id="AAC38064.1"  
/db\_xref="GI:2624951"

/translation="MANEDFELHYLKKAAADLRDARQIQLEAREPREIAIVSMACR  
LPGVDSPSLMDLVAAGSDGVSAFPIDRGMDLEGIYDDPDHPGTTYARDGGFLDDA  
AGDPFAFGISPREALAMDPOORLLLETTWEVLERAGIDPAALRGSRGTGVTGVIEND  
YAWSADDTDLQLEGYLVNGSAAVASGRVSVTFEGEPVSLDTACSSLSVALHVAIQ  
ALRSGECLTITGVTVLASVAFPIEFBSQGLADDCRCKAFADANGFMAEGIGLL  
LVERLSDAQRLGHEVLAVVRGAVNQDASNGLTAFPTGFSQQRVIRDAITDARLTVADV  
IDILEAHGTGRLGDEIEAQALIAITYGQDRQPVWLGSKNIGHTQAAAAGVAGVKT  
VAMRHGIMPRTLHIDETPEVDVTPVWLGSLKNSNIGHTQAAAAGVAGVAGVKT  
NAHVILEQAPAEPEVDVPTPAGVPWVLGAHTGPALSALAEALAAHEVDEAPLGDA  
SLHGAARFSCRAVMIGSTRDPFAELREALTCGSRASKRVVFPQGGQSMWIGRGL  
MESSAVERDSVILRTDRAAEFVDMASAAVLAGEPTPDLDRVDVQVPLFTVMVALAN  
VWRSYGVPEAAVVGHSQGEITAAHVAAGLSLRDAARVVALRSRALTALAGEGGMVVQ

Query Match 4.8%; Score 24; DB 1; Length 32870;  
Best Local Similarity 100.0%; Pred. No. 0.92; Mismatches 0; Indels 0; Gaps 0;  
Matches 24; Conservative 0

QY 429 CTACGGCAGCACCGACGACCTGCGG 452  
|||||  
DB 2383 CTACGGCAGCACCGACGACCTGCGG 2406  
|||||

RESULT 8  
AP002998  
LOCUS 347750 bp DNA linear BCT 15-MAY-2001  
DEFINITION Mesorhizobium loti DNA, complete genome, section 5/21.  
ACCESSION AP002998 BA000012  
VERSION AP002998.2 GI:14022369  
KEYWORDS  
SOURCE Mesorhizobium loti  
ORGANISM Mesorhizobium loti  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Phyllobacteriaceae; Mesorhizobium.  
REFERENCE  
1 (sites)  
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,  
Watanabe, A., Iidesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,  
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,  
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,  
Takeuchi, C., Yamada, M. and Tabata, S.  
Complete genome structure of the nitrogen-fixing symbiotic  
bacterium Mesorhizobium loti  
DNA Res. 7 (6), 331-338 (2000)  
JOURNAL  
MEDLINE 21082930  
PUBMED 11214968  
REFERENCE 2 (bases 1 to 347750)  
Kaneko, T.  
Direct Submission  
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research, Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail:kaneko@kazusa.or.jp)  
URL: http://www.kazusa.or.jp/rhizobase/  
Tel: 81-438-52-3935(ex.2338), Fax: 81-438-52-3934)  
On May 11, 2001 this sequence version replaced gi:11994973.  
FEATURES  
source  
1..347750  
/organism="Mesorhizobium loti"  
/mol\_type="genomic DNA"  
/strain="WAF303099"  
/db\_xref="taxon:381"  
/complement (12..2333)  
/gene="ml11656"  
/complement (12..2333)  
/gene="ml11656"  
/codon\_start=1  
/transl\_table=11

```

/product="ATP-binding protein of dipeptide ABC
transporter"
/protein_id="BAB48979.1"
/db_xref="GI:14022370"
/translation="MNEAVRNPAPEIIEIENLSISFFTRKGEIPAVMDFSCTVMPGEA
MGIYVESGCKSTVSLGIMDLNLSIKGIVGRIRKFOGKDMGELSEELRAIRGNKIAM
IYQPMASLNPAMKVGQOLMEVPLIHDKVSKEAVRSLKLEWRAVKLPDPERMWSYP
HOLSGQOORLVIAWALLSKPALLDEPTTALDVTVEAGIVELVKGIGKFGFTSMIF
VSHNLGLILETCDRIITWYSGEAVETGKIADVDFMRHPYTGQILFRSIPLPQADKNSR
PLIIPGQPLPHRPGKCNFGPRCHHEFVEGVCNAAEIPMVEVAGHGFHFKRCVRFNE
IDWALPFGAKVNERVVPVGLIKIEDIRKYIKVGVSEGVFGSGSEGRVYKANETISEM
ARESETVAIVGESGCKSTLAKVILGLETSAGVTILGNKOIQSTGIESRVEYVSSI
QWVFNQDPLNPHSGVSOIIRLTLEKNVCKTVADRRKRLMLELIDLVKLPRAETFKR
PROLGGKQRIQVAVRAGAKVAVDEPVSALDVSQAAVETLMDIQRKNTMML
FISHDLSVVRVARIUVVMYLYIVEQGTQIQIEFAPHPYTEALLSALPIADTSYVK
RHIIVLEGDIPSAMNPPGCTPQTRCGYKKYKVDNICETKPPVVKHGLDGHGSLCWLD
DVLAKMEFVIRKDEHAAHEGVDPDPAHGTGTFGFGFSGPPARPSPGKNNAEADATGKAV
EDTDAVLRARHREVRDETSETGVSKPKDTRRH"
/complement(2382..3539)
/gene="ml11657"
/complement(2382..3539)
/gene="ml11657"
/codon_start=1
/transl_table=11
/product="permease protein of dipeptide ABC transporter"
/protein_id="BAB48980.1"
/db_xref="GI:14022371"
/translation="MQVQIIISAFITVILEVLSRFPWFWPWTALIIVMGASFYKKKLALYG
QLFSGVGIYGVIGICLFWLFTAFIFASTISPFDPQLAQVPMKMDLPGLTEPOSGLYLF
GLDKLARVDFRVMYVSGQITLIAPAATGAFALMGVITGLPAGYTGKIDTIVLSFIAN
LVLAPEITLLFYLLYPTGMDPTPIYAMAGVFPFPIIFFSVFWYRKNRDPRIYIL
LGLTILIGWVILGVAFDRDPLHIVHDNPOLNIKVAVVFASSPGVFRIVRGLWMDIK
TRDYVAAQTRGESFWYIMWELLNARGPLIVDACILRGYITILLGLTGLVGLGLAP
ESPDMTAKDRASRLRSLFIHPALPPTIALMSVFLGLNLLADLSREQSKND"
/complement(3552..4586)
/gene="ml11659"
/complement(3552..4586)
/gene="ml11659"
/codon_start=1
/transl_table=11
/product="permease protein of oligopeptide ABC
transporter"
/protein_id="BAB48981.1"
/db_xref="GI:14022372"
/translation="MLSFVIRLGTWALTMCLTMTWVFFLINLBNLKLALISQTEMH
TSAEQLESWLVNHGRQNFVFRYQWLGVMKQPVTDTEATGPAQRFSCNDPVVPTF
AGVVEGDFGCKTKFTVASKLFPALGATGILMFCVLAVMVPVSLIIGILAGMRGSR
TDRVLSVASTATPEYVSGVIFTVIFASWLGWLGSAASQGITFYNFPLPVMTL
ATYIGIYIARMTASRVMTQOYIIRTKLGLSGVGVVVKHARNALAPFTVIMLQ
FPMLLTGVIIVEMFYQGFYTLVEAAGNNDIDLLGCSLYSVFVILTITQLSDVGY
AFLNPRIRVQ"
/complement(4809..6464)
/gene="ml11660"
/complement(4809..6464)
/gene="ml11660"
/codon_start=1
/transl_table=11
/product="dipeptide binding protein component of ABC
transporter"
/protein_id="BAB48982.1"
/db_xref="GI:14022373"
/translation="MTIYKSGDRVPDHLMAAEHAKGVDRERFLALASVFGASTA
MAYGLLGLAPTEAREADAPKGGIILKVAQWIKPDPRKSDWSEIANERQALFLIV
KTYTETFRPYLLESWVNDADQYVLVHRKGVTVHNGDAFTADDVWNFNWGEKALV
EGNSMPRLGTLFDKTKLREGAVTKDDMTVKLSLPEDIISIIPGICDYVALIVHK
SFDEKGDFFKACPITGTPFELSYDVGVKVKYRADNKWDGDVYLDGIEFTDYGTD
PAAWVSAPAGEVHTFTFSADYVSLDKMDLVKRSVWTAITVCTVNTVTHQPNDEK
VRKALQYDNNVVLQIYGNAGSVAENHVCISLHPEYVELPKLTFDPAKAKALMAEA
GQADFHELITVDWEDHKNTGDAIAQLRDAGIKVKRTVILPGSTFNDWMLKPLSWTN
WNNRPLGVQVLYATYEGTAWNECWSNPDNDKNAALAVSDAAKRKEMKDIQIIL
QDSGIITQYWRKLYSHKVAQVQNYKHPTFEDRYGKIWLDDQA"
/complement(6717..12953)
/gene="ml11661"

```

```

/complement(6717..12953)
/gene="ml11661"
/translation="MPVRCRRSRFTSPTSRSAPASACATRRRSARCASMRPFRSTAI
PTIRILASMPASGRSDMKTVMRILRIVLVTLIVAGAILVLTTERGRDLNLAG
IISRLASDDRRKYTVSGDGIWSGALSVHDVHLEDRQCAWLVARVWDSPELALLSK
SFSADRIIAERIELARLPVVTGPPQAGGTITLTPVSIDIREIDELPEALGQOLASGSI
AELAKMKLKAADAPLAVETVLTNRHDKQGNDAKIFAPADRDLDLDKFKMLGTTLA
GIITANLLKLPDTPVDIVASGTGPLANWNGIATFVVDGKIVTQITGRQLDITKGNHVE
AKGDEPAPFLPDNLSLFAKGTDFDVAGTATAGGISIDRASIESDAIHGTATAD
PAGASDLIVETAAKSAPQIVVGSAAQPIITVAIRNATARAFDGGKAPITVDIGASLASV
VAGTQNLNDISAQIHSDFDQNRAGPVAVKLTAGGLTDVATLPLATVKGVDVLGAG
VLSKDAITADISQNLNALSNTVALDLSKMLKNWADAVSTALPQQISVILGER
VKFSAAATRDPPQGAFAANSISLTSGLSASGTASAGSDIQADIKGTIGDVSLSVMV
GVPVAGVNFALSGAGTAPDFTVSADSDSLTAGSRTVXAIKLTASGADITASPSAD
VSLTGSVNDQPLDLKAALVTSEGRKRSINGFLVSLGDNKVSGLDLDKFKMLGTTLA
APADOLAALAGQAVTGDTINGTILPAKGDAPSVTIDAKSTSILARGEVTAIAVNAL
IANYLKAPASISGTVKADSVTSGATEISIGIVDLKEDGDMTNGTGTAGIAPATAAGR
VKIAGQTTVSVEIASGEATVRGKAIAEPSTIITANGTTSIEKLALDVGGSVTLTSGS
AGQTLDLAAQFSALPAALANDFAPGLDAVTLGGAAQTVTGPSPEVRFNAQLGDLT
SQTQAEIPLGLVADAAGVYITLAGVVALDHALTSLGDKISGTASGTTNPGASFDALDT
SSQPSLPLTVGSAASPVRIEIRSLSAKVAGESTRARLDVSAIILPITTSIPAKVDGLVL
ALHSDAPLTKNRAGPVSGTVSDVKVLONAAIAPLIAGKVTAAALNARITSDGVALDSG
SLKSDALNSQVAGQVSLRDGALDLKLADAPSSALPAAARGMLGORAQISATLKRDPN
GGINVDGLVGGTLAGTAAQGLADNKLADLTGALADVSLPSKDKAGAIAPALNAQG
PAIAPADSLTIVNSRLLVAEREITGLKLTATGKADAANPAANVQITGVAGQALOGSA
VLATADGKSAINGLLISLGRKISGDLADLPKFEPTGISLPLDIPGLAALAEKAE
GVDRTIATFTKNGAAPVAIKAAATASITRGLOKATVSDIALIANVLAAPVTSKIRA
DVTVSGGVTISGIDVDLARDGDTGFGSGATVKGIPAKAAGRVKANGVTTIELASGO
ATVQIKAAIAQAQSVTSIANGTITLDRLVNLGGTATVTVGKQALDINAVLKVPM
SLANGFSPGLDAAGSISGTVKVTGAPSPAFVAFNIDAGVQTSIRGAGVAVSVSS
GTFAGNKLTFNANIDGGLKGGTITTAGTPALVDLNFVAGPMLUSQKLAAGL
SLSGTANVNVQVRGPATSPVIGGVSSTSGARLVARSGLVADLADVSIIGGVARIN
RLTGTLSTRGSLASGTVGINPAQGFADLSIKLTDGRVTRGVTVANLGGDLITKGP
LVSPVIAGTVNLAKTVITVDPKLPASISALVDKHNKNAPEAVRAQDKALRPARTSGKS
GSGSLADVTYNAPNQIIFIQGRGVDAELGGSLRLTGPASSPOAVGTFTLQGRGLSILG
KRLTFTGTVFGVSGSLPYLNLATTTTTTSATVTVSVSGEATNPKFTFSVPALPEDE
VLAQILFGRSMNSLPLQIAQLAAEAGQLAGVGSSTSLLENLSRISAIGVDLDDVTTDDK
GGTAVSAGKYLNDRTVYTIQKDKFPGSGKATIDLVNVRGVKLRGEATDAGEAKGGIFY
EKEY"
/complement(12772..14625)
/gene="ml11662"
/complement(12772..14625)
/gene="ml11662"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation="MLCATALVAESRPAAPAFELFGIKLWSSNDEADADIVDPLRYAVT
ITAPADKDLVKRLNLSALKGDEDRPVSGSLGLMAKARSDEQLVAALFADARYEGV
VITITDGLDELPPDAEPKQPIPVITINIAVGPKFTLGTIRLEGEAAGLMSADYGL
TISGAGSAGVLKAEALIVRTIKEOGRPLAKVTDRIIVADHATSLTDVTLTVAAAGPVA
GYCATVTEGTEKVRDFTFYMVTGLRGKQYSQEIISDARDRLALAEVNSVTFKEADK
LDADGNIPIGVQVVSERKPRYFGGTFNTEGLLEGYWGHRNLFPGHAEKRLIDGALS
GIGSNMLSDNTNAGIMPEKPGVIGPASKFPAFTVLEHPDAYDHFHFKSGSTGLSYE
LDKQOTVSAEALDYSLTDAFKHTYLIASVPLQVYVNDNRSLNPTGRFGRFLYAYE
PSYDITMSGAFLKAGESAYOSLDTASKVLAERVAIGSIIVGTQLQNPADRRFYVSG
GGSGVRGVAGYIGIKPDKFTGQPIGGLSPETSVEMRIAVTDTIGIIVFPVDAGVSTKS
VNFSDVKVAGVGLRYVTFPGPLIDAAVPLNRDNDPHFGIYAGIGQAF"
/14972..20458
/gene="ml11663"
/14972..20458
/gene="ml11663"
/translation
```

/translation="MAMRAARGSLIILLFFALWNGVAQAEARRIATDSDYRGF  
DLBQDQNTLDCKTCLGDPACRATYNTAKWCLKSDYNOLKEPNGAVAGVKVI  
DGDFFLIGAPPELLFFPNWADQAOQVNRLLGPAFKPTEGMAILLSSRAEQVLTGDH  
RSAMHYEAVSVMPDQGLWMLARETLAVQPAINTSEASTUPMNAISAAFNAYKLL  
RTTKTRADVLALGAGLDKRDLYRPAQAYEASLAVTSPAVQADYADLKARGFVRV  
DHSVDATSGAPRICAFSEBELVKTGVYAOFTVDNAPKGVKAEKQICVEGLEHGQ  
HYDVTFRAGLPAAGVTRAPVVLSTVQDRAPSAFTGDSFVLPAGARRGIPVWTWN  
MNAEMKLYEIGRSLAQLLSGVQFLHOLDSDISNISEOMGAPVQMGKLDIYNDLKN  
EVTTSFVDRALQKRGVYVILTRQAVDQKDDYNSLATONFVVSIGLSTYTGQDGL  
NVFARSUGSAKRGVYVILTRQAVDQKDDYNSLATONFVVSIGLSTYTGQDGL  
AQGDNDVFLDMSKAGFDLSDRGVTRGAAPGALDYAVTERTGIRAGEDVHVAALAR  
DGAAXAVENLPLATIFSRPDGVEDRIVSDGASAGGHAVELPLEPNMRGTWSVSIHT

Query Match 4.8; Score 24; DB 1; Length 347750;  
Best Local Similarity 100.0%; Pred. No. 0.71; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 468 CCGCGCGGACGGCGCGCGTGA 491  
|||||  
Db 22495 CCGCGCGGACGGCGCGCGTGA 22518  
|||||

RESULT 9  
MLCB1779 43254 bp DNA linear BCT 08-AUG-1997  
LOCUS Mycobacterium leprae cosmid B1779.  
DEFINITION  
ACCESSION Z98271  
VERSION Z98271.1 GI:2326678  
KEYWORDS cell division ATP-binding protein; ferredoxin; fold; ftsE;  
homoserine O-acetyltransferase; methylenetetrahydrofolate  
dehydrogenase; monophosphatase; peptide chain release factor 2;  
prfB; pseudogene; small protein B; smpB; transfer RNA-Met; trpS;  
tryptophanyl-tRNA synthetase.

SOURCE  
ORGANISM Mycobacterium leprae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
1 (bases 1 to 43254)  
Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.  
Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae  
Mol. Microbiol. 7 (2), 197-206 (1993)  
93188700  
8446027  
2 (bases 1 to 43254)  
Hamlin, N. and Churcher, C.M.  
Unpublished  
Direct Submission  
Submitted (07-AUG-1997) Mycobacterium leprae sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge,  
CB10 1SA E-mail: barrall@sanger.ac.uk Cosmids supplied by Dr.  
Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,  
Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,  
France Requests for cosmids should be sent to Karin Eiglmeier  
(kei@pasteur.fr)

Notes:  
The Sanger Centre is funded to complete the sequence of M. leprae  
by the Heiser Program for Research in Leprosy and Tuberculosis of  
The New York Community Trust.  
Work in Paris is supported by the Heiser Trust, the Association  
Francaise Raoul Follereau and the Groupement de Recherches et des  
Etudes des Genomes (GIP-GRG).  
Details of M. leprae sequencing at the Sanger Centre are available  
on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/>)  
CDS are numbered using the following system eg MLCB33.01c. ML (M.  
leprae), cB33 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE

COMMENT  
The Sanger Centre is funded to complete the sequence of M. leprae  
by the Heiser Program for Research in Leprosy and Tuberculosis of  
The New York Community Trust.  
Work in Paris is supported by the Heiser Trust, the Association  
Francaise Raoul Follereau and the Groupement de Recherches et des  
Etudes des Genomes (GIP-GRG).  
Details of M. leprae sequencing at the Sanger Centre are available  
on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/>)  
CDS are numbered using the following system eg MLCB33.01c. ML (M.  
leprae), cB33 (cosmid name), .01 (first CDS), c (complementary  
strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been analysed. Gene prediction is based on  
over 100 codons have been analysed. Gene prediction is based on  
positional base preference in codons especially where there is an  
increase in the observed/expected third position G + C. CAUTION:  
We may not have predicted the correct initiation codon. Where  
possible we choose an initiation codon (atg, gtg, or ttg) which is  
preceded by an upstream ribosome binding site sequence (optimally  
5-13bp before the initiation codon). If this cannot be identified  
we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions. Cosmid B1779 is  
overlapped at the 5' end by EMBL:ML022 M. leprae cosmid L308. Note  
that there are 11 sequence conflicts between this sequence and  
EMBL:ML022; in each case our sequence has been checked and is  
thought to be correct.

FEATURES  
source  
1..43254  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1769"  
/clone="cosmid B1779"  
misc\_feature  
1..4383  
/note="overlap with EMBL:ML022 M. leprae cosmid L308 from:  
32036 to: 36410"  
complement(1..239)  
/gene="trpS"  
complement(<1..239)  
/gene="trpS"  
/note="MLCB1779.01c, trpS, tryptophanyl-tRNA synthetase,  
partial CDS, len: >79 aa; similar to many eg. SVW HAEIN  
P43835 tryptophanyl-tRNA synthetase (ec 6.1.1.2) [334  
aa] initn: 205 initl: 134 opt: 227 z-score: 409.4 E():  
7.4e-168 Smith-Waterman score: 227; 55.1% identity in 69 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="tryptophanyl-tRNA synthetase"  
/protein\_id="CAB10982.1"  
/db\_xref="GI:2326679"  
/translation="MSTATGFSRIFSGVQPTSDSLHGNALGALTQWVALQYDHPDEY  
EAFECVVDLHAIITIAQDPETILWRRLTVAQAIALG"  
308..1186  
/gene="MLCB1779.02"  
308..1186  
/gene="MLCB1779.02"  
/note="MLCB1779.02, unknown, len: 292 aa; some similarity  
to TR:055217 (EMBL:L35154) esterase daup from streptomycetes  
sp. (298 aa), fasta scores; opt: 138 z-score: 253.6 E():  
3.5e-07, 26.9% identity in 204 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein MLCB1779.02"  
/protein\_id="CAB10983.1"  
/db\_xref="GI:2326680"  
/db\_xref="GOA:O32873"  
/db\_xref="SPTREMBL:O32873"  
/translation="MVACGTRGTTFRVSGMRQPIHLGSGEVLHLLPMMQSVWEIY  
APQANVERVEVFAPVMAAHNGPSSASSMSTLADHVEQQLDELGWDTHLVGNS  
LGVWVTELEBERGRARSVTGTAPAGWTRWSLAKFVIAKVGAPLIVAVAGLGSRA  
LSLPSHRLASLPLTGSFAGLSEHQLNIVEDAGHCPCAYFLLIKTLIQGLRELADI  
FIVLVLLCEAKDQVLPPEPKYSRHRFDHLPRTQVTKDGMGHVPMFAPEVITKLITN  
FINQCSKPAEASPAS"  
complement(1183..1446)  
/note="MLCB1779.03c, probable isocitrate dehydrogenase  
(nadh) pseudogene fragment, len: 87 aa; similar to part of  
many eg. IDH\_SPHYA P50215 isocitrate dehydrogenase (nadh)  
(406 aa) (43.4% identity in 83 aa overlap)"  
/pseudo

FEATURES  
source  
1..43254  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1769"  
/clone="cosmid B1779"  
misc\_feature  
1..4383  
/note="overlap with EMBL:ML022 M. leprae cosmid L308 from:  
32036 to: 36410"  
complement(1..239)  
/gene="trpS"  
complement(<1..239)  
/gene="trpS"  
/note="MLCB1779.01c, trpS, tryptophanyl-tRNA synthetase,  
partial CDS, len: >79 aa; similar to many eg. SVW HAEIN  
P43835 tryptophanyl-tRNA synthetase (ec 6.1.1.2) [334  
aa] initn: 205 initl: 134 opt: 227 z-score: 409.4 E():  
7.4e-168 Smith-Waterman score: 227; 55.1% identity in 69 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="tryptophanyl-tRNA synthetase"  
/protein\_id="CAB10982.1"  
/db\_xref="GI:2326679"  
/translation="MSTATGFSRIFSGVQPTSDSLHGNALGALTQWVALQYDHPDEY  
EAFECVVDLHAIITIAQDPETILWRRLTVAQAIALG"  
308..1186  
/gene="MLCB1779.02"  
308..1186  
/gene="MLCB1779.02"  
/note="MLCB1779.02, unknown, len: 292 aa; some similarity  
to TR:055217 (EMBL:L35154) esterase daup from streptomycetes  
sp. (298 aa), fasta scores; opt: 138 z-score: 253.6 E():  
3.5e-07, 26.9% identity in 204 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein MLCB1779.02"  
/protein\_id="CAB10983.1"  
/db\_xref="GI:2326680"  
/db\_xref="GOA:O32873"  
/db\_xref="SPTREMBL:O32873"  
/translation="MVACGTRGTTFRVSGMRQPIHLGSGEVLHLLPMMQSVWEIY  
APQANVERVEVFAPVMAAHNGPSSASSMSTLADHVEQQLDELGWDTHLVGNS  
LGVWVTELEBERGRARSVTGTAPAGWTRWSLAKFVIAKVGAPLIVAVAGLGSRA  
LSLPSHRLASLPLTGSFAGLSEHQLNIVEDAGHCPCAYFLLIKTLIQGLRELADI  
FIVLVLLCEAKDQVLPPEPKYSRHRFDHLPRTQVTKDGMGHVPMFAPEVITKLITN  
FINQCSKPAEASPAS"  
complement(1183..1446)  
/note="MLCB1779.03c, probable isocitrate dehydrogenase  
(nadh) pseudogene fragment, len: 87 aa; similar to part of  
many eg. IDH\_SPHYA P50215 isocitrate dehydrogenase (nadh)  
(406 aa) (43.4% identity in 83 aa overlap)"  
/pseudo

FEATURES  
source  
1..43254  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1769"  
/clone="cosmid B1779"  
misc\_feature  
1..4383  
/note="overlap with EMBL:ML022 M. leprae cosmid L308 from:  
32036 to: 36410"  
complement(1..239)  
/gene="trpS"  
complement(<1..239)  
/gene="trpS"  
/note="MLCB1779.01c, trpS, tryptophanyl-tRNA synthetase,  
partial CDS, len: >79 aa; similar to many eg. SVW HAEIN  
P43835 tryptophanyl-tRNA synthetase (ec 6.1.1.2) [334  
aa] initn: 205 initl: 134 opt: 227 z-score: 409.4 E():  
7.4e-168 Smith-Waterman score: 227; 55.1% identity in 69 aa  
overlap"  
/codon\_start=1  
/transl\_table=11  
/product="tryptophanyl-tRNA synthetase"  
/protein\_id="CAB10982.1"  
/db\_xref="GI:2326679"  
/translation="MSTATGFSRIFSGVQPTSDSLHGNALGALTQWVALQYDHPDEY  
EAFECVVDLHAIITIAQDPETILWRRLTVAQAIALG"  
308..1186  
/gene="MLCB1779.02"  
308..1186  
/gene="MLCB1779.02"  
/note="MLCB1779.02, unknown, len: 292 aa; some similarity  
to TR:055217 (EMBL:L35154) esterase daup from streptomycetes  
sp. (298 aa), fasta scores; opt: 138 z-score: 253.6 E():  
3.5e-07, 26.9% identity in 204 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein MLCB1779.02"  
/protein\_id="CAB10983.1"  
/db\_xref="GI:2326680"  
/db\_xref="GOA:O32873"  
/db\_xref="SPTREMBL:O32873"  
/translation="MVACGTRGTTFRVSGMRQPIHLGSGEVLHLLPMMQSVWEIY  
APQANVERVEVFAPVMAAHNGPSSASSMSTLADHVEQQLDELGWDTHLVGNS  
LGVWVTELEBERGRARSVTGTAPAGWTRWSLAKFVIAKVGAPLIVAVAGLGSRA  
LSLPSHRLASLPLTGSFAGLSEHQLNIVEDAGHCPCAYFLLIKTLIQGLRELADI  
FIVLVLLCEAKDQVLPPEPKYSRHRFDHLPRTQVTKDGMGHVPMFAPEVITKLITN  
FINQCSKPAEASPAS"  
complement(1183..1446)  
/note="MLCB1779.03c, probable isocitrate dehydrogenase  
(nadh) pseudogene fragment, len: 87 aa; similar to part of  
many eg. IDH\_SPHYA P50215 isocitrate dehydrogenase (nadh)  
(406 aa) (43.4% identity in 83 aa overlap)"  
/pseudo

```

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10984.1"
/db_xref="REMTREMBL:CAB10984"
complement(1542..1682)
/notes="MLCB1779.04c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 46 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (56.5% identity in 46 aa overlap)"
/pseudo

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10985.1"
/db_xref="REMTREMBL:CAB10985"
complement(1684..1941)
/notes="MLCB1779.05c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 85 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (63.6% identity in 33 aa overlap)"
/pseudo

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10986.1"
/db_xref="REMTREMBL:CAB10986"
complement(1990..2106)
/notes="MLCB1779.06c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 38 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (61.1% identity in 36 aa overlap)"
/pseudo

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10987.1"
/db_xref="REMTREMBL:CAB10987"
complement(2107..2295)
/notes="MLCB1779.07c, probable isocitrate dehydrogenase
(nadp) pseudogene fragment, len: 62 aa; similar to part of
many eg. IDH SPHYA P50215 isocitrate dehydrogenase (nadp)
(406 aa) (56.5% identity in 62 aa overlap)"
/pseudo

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10988.1"
/db_xref="REMTREMBL:CAB10988"
2820..3149
/notes="MLCB1779.08, probable pseudogene fragment, len: 109
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (33.7% identity
in 89 aa overlap)"
/pseudo

misc_feature
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10989.1"
/db_xref="REMTREMBL:CAB10989"
3083
/notes="conflict: this base is missing in EMBL:ML022 M.
leprae cosmid L308"

misc_feature
3208..3210
/notes="conflict: ACG is GC in EMBL:ML022 M. leprae cosmid
L308"

misc_feature
3242
/notes="conflict: this base is missing in EMBL:ML022 M.
leprae cosmid L308"

CDS
3429..3782
/notes="MLCB1779.09, probable pseudogene fragment, len: 141
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (37.7% identity
in 130 aa overlap)"
/pseudo

CDS
/codon_start=1
/transl_table=11
/db_xref="PSEUDO:CAB10990.1"
/db_xref="REMTREMBL:CAB10990"

```

```

CDS
3767..4096
/notes="MLCB1779.10, probable pseudogene fragment, len: 109
aa; similar to part of MT17 YEAST P06106
o-acetylhomoserine sulphydrylase (443 aa) (36.0% identity
in 111 aa overlap)"

Query Match 4.6%; Score 23; DB 1; Length 43254;
Best Local Similarity 100.0%; Pred. NO. 2.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GCCGACGGCAGCTCCGCCGACGA 230
|||||
DB 21891 GCCGACGGCAGCTCCGCCGACGA 21913

RESULT 10
MLEPRTN3 312050 bp DNA linear BCT 20-FEB-2001
LOCUS Mycobacterium leprae strain TN complete genome; segment 3/10.
DEFINITION AL583919 AL450380
ACCESSION AL583919.1 GI:13092766
VERSION
KEYWORDS
SOURCE Mycobacterium leprae
ORGANISM Mycobacterium leprae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE 1 (bases 1 to 312050)
Cole, S.T., Bigmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,
Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,
Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,
Taylor, K., Whithhead, S., Woodward, J.R. and Barrell, B.G.
Massive gene decay in the leprosy bacillus
NATURE 409 (6823), 1007-1011 (2001)
JOURNAL MEDLINE
PUBMED 21128732
11234002
REFERENCE 2 (bases 1 to 312050)
Parkhill, J.
Direct Submission
Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
Notes:
Details of M. leprae sequencing at the Sanger Centre are available
from http://www.sanger.ac.uk/Projects/M_leprae/ A relational
database containing the M. leprae sequences is available from
http://genolist.pasteur.fr/leprona/.
FEATURES
source
1..312050
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/strain="TN"
/db_xref="taxon:1769"
complement(2412..3665)
/notes="LEPRTN, copy 1, 1254 bp. Shows 99-100% identity
with regions 72707..728460, 1539937..1541188,
1480312..1481018, 2203983..2204515, 691673..691731"
complement(2413..2473)
/notes="61 bp repeat found at 3'-end of LEPRTN and 5' to
REPUPP"
gene complement(5611..6426)
/gene="ML0569"
complement(5611..6426)
/gene="ML0569"
/notes="Similar to M. tuberculosis Rv1433, putative
membrane protein, TR:006825 (EMBL:AL123456) (271 aa);
Fasta score E(): 0, 68.3% identity in 271 aa overlap."
CDS

```

Shows weak similarity to other hypothetical proteins in M. tuberculosis. Previously sequenced as TR:Q49706 (EMBL:U00013) (271 aa); Fasta score E(): 0, 100.0% identity in 271 aa overlap.

Similar to ML0426, ML2446 and ML2664

/codon start=1  
/transl table=11  
/product="putative membrane protein"  
/protein\_id="CAC30077.1"  
/db\_xref="GI:13092768"  
/db\_xref="SPTREMBL:Q49706"  
/translation="MGIMRAAIRGLVVLVIGIAATVIAQVRVSLAAVNHAGFAITSI LPTDAVGVAPVVTFTGAPVTRKATELSLKVRSAFAMNGKFDMLDSKVQVWVDR YWPESTTALTVGNLSNFTKGTGAILADIASHFTFTVTDGVEAETPPPLSPHRRP HWRGEGVWASMGKTEFPPTGKTVMSKORSVIMDSVSGIPVDDPEGYRLSDVYAV RISSGLGIHSPAWQSMGFNTSGHGLSLSPADAENYNTVNIQDPVIVKE"  
6876. .6882  
/note="possible RBS"  
6889. .7908  
/gene="gap"  
/note="synonym: ML0570"  
6889. .7908  
/gene="gap"  
/EC number="1.2.1.12"  
/note="Similar to M. tuberculosis gap, Rv1436, glyceraldehyde 3-phosphate dehydrogenase, SW:G3P MYCTU (006822) (339 aa); Fasta score E(): 0, 89.1% identity in 339 aa overlap. Similar to many e.g. Mycobacterium avium gap, glyceraldehyde 3-phosphate dehydrogenase, SW:G3P MYCAV (P94915) (339 aa); Fasta score E(): 0, 87.9% identity in 339 aa overlap. Previously sequenced as SW:G3P MYCLE (P46713) (339 aa); Fasta score E(): 0, 99.7% identity in 339 aa overlap. Contains Pfam match to entry PF00044 gpdh, glyceraldehyde 3-phosphate dehydrogenases. Contains PS0071 Glyceraldehyde 3-phosphate dehydrogenase active site."  
/codon start=1  
/transl table=11  
/product="glyceraldehyde 3-phosphate dehydrogenase"  
/protein\_id="CAC30078.1"  
/db\_xref="GI:13092768"  
/db\_xref="GOA:P46713"  
/db\_xref="SWISS-PROT:P46713"  
/translation="MTVRVINGFGFRIGRNFYRALLAQOEHGIADVQVVAINDITDNS TLAVLKFDLSILGRPHDVSLEEDTIVGSEKIKALVREGPAALPWHAFGVGVVVE STGLFTVAAKAKHLEAGAKKIVLSAPATDPDITIVGVNDKVDGSGNIIISASCTT NCLAPLAKVLHDQFGIVGLMTVTHAYTQDNLODGHSLRPARAALNVPTSTGA AKAIGLWPELKGIDGIALRVPIPTGVSVDLADLSKVSVEINAVFQDAEGRKIL GILKYVDATVSSDITVDPHSSIFDSGLTKVIASQAKVSWSDNMGWSNRLVDLVGL VGKSL"  
6892. .7902  
/gene="gap"  
/note="Pfam match to entry PF00044 gpdh, glyceraldehyde 3-phosphate dehydrogenases, score 733.80, E-value 7.9e-219"  
7354. .7377  
/gene="gap"  
/note="PS00071 Glyceraldehyde 3-phosphate dehydrogenase active site"  
7911. .9161  
/gene="pgk"  
/note="synonym: ML0571"  
7911. .9161  
/gene="pgk"  
/EC number="2.7.2.3"  
/note="Similar to M. tuberculosis pgk, Rv1437, phosphoglycerate kinase, SW:PGK MYCTU (006821) (412 aa); Fasta score E(): 0, 80.6% identity in 412 aa overlap. Similar to many e.g. Corynebacterium glutamicum pgk, phosphoglycerate kinase, SW:PGK CORGL (Q01655) (403 aa); Fasta score E(): 0, 59.2% identity in 407 aa overlap. Previously sequenced as SW:PGK MYCLE (P46712) (416 aa); Fasta score E(): 0, 99.8% identity in 416 aa overlap.

Contains Pfam match to entry PF00162 PGK, Phosphoglycerate kinases. Contains PS00111 Phosphoglycerate kinase signature."

/codon start=1  
/transl table=11  
/product="phosphoglycerate kinase"  
/protein\_id="CAC30079.1"  
/db\_xref="GI:13092769"  
/db\_xref="GOA:P46712"  
/db\_xref="SWISS-PROT:P46712"  
/translation="WRIPNLKDLLEEGVGRCLVRCNDLVFLGDDGAIITDLGRVTAS VPTKALAEAGAKTVAAHGRKPNKPKLSLEPAAALGEOLQCNQVLVCSTDRSP VGTALACVERLTDCDILLQNIREDPETSVDDEIRALAKOLVELVGSAGFVSDG FGVHRRQASVIDVAITLPHAGILVADEIRILEOLTSAPKPYAVLVGSGVSKLIG VISTEKEVADSPQVTAADAIPTDGLMDIGPESVKRFATILSNASTITVWNGPVGVS EFPAVACTRGVAAEIAVATGKAFSVVGGSDSAAAMRALSLPEGSVSHLSTGGGASL EYLEGKTLPGIEVLGREQPRGDS"  
7911. .9140  
/gene="pgk"  
/note="Pfam match to entry PF00162 PGK, Phosphoglycerate kinases, score 757.00, E-value 7.9e-224"  
7962. .7994  
/gene="pgk"  
/note="PS00111 Phosphoglycerate kinase signature"  
9149. .9153  
/gene="pgk"  
/note="possible RBS"  
9158. .9943  
/gene="tpi"  
/note="synonym: ML0572"  
9158. .9943  
/gene="tpi"  
/EC number="5.3.1.1"  
/note="Similar to M. tuberculosis tpi, Rv1438, triosephosphate isomerase, SW:TPIS MYCTU (008408) (261 aa); Fasta score E(): 0, 83.9% identity in 261 aa overlap. Similar to many e.g. Corynebacterium glutamicum tpi, triosephosphate isomerase, SW:TPIS CORGL (P19583) (259 aa); Fasta score E(): 0, 64.5% identity in 256 aa overlap. Previously sequenced as SW:TPIS MYCLE (P46711) (261 aa); Fasta score E(): 0, 99.6% identity in 261 aa overlap. Contains Pfam match to entry PF00121 TIM, Triosephosphate isomerase active site."  
/codon start=1  
/transl table=11  
/product="triosephosphate isomerase"  
/protein\_id="CAC30080.1"  
/db\_xref="GI:13092770"  
/db\_xref="GOA:P46711"  
/db\_xref="SWISS-PROT:P46711"  
/translation="MSRKSLIAGNWKMNLMHFEALVQKIAFSLPDKYKVDVTVVL PPTDLSRVQTLVDGKLRTYGAOLDSQHDIGATDGSAGFLAKIGCSFVLVGHSE RRYIHDEGALVAAKTAALKNSLTPTVICIYEIVREVGHVSHQCNQRLRGSILAGLS PRQIGNVIVIEFWALITGRVASADAQEVCEAIRKELGALASQVAETVRVLYGGS LNAKNTGDIVAQDDVDGGLVGGASLDGAQATLAVIAGGFLP"  
9164. .9931  
/gene="tpi"  
/note="Pfam match to entry PF00121 TIM, Triosephosphate isomerase, score 524.60, E-value 3.7e-169"  
9665. .9697  
/gene="tpi"  
/note="PS00171 Triosephosphate isomerase active site"  
9927. .10202  
/gene="ML0573"  
9927. .10202  
/gene="ML0573"  
/note="Unknown function. Previously sequenced as TR:Q49684 (EMBL:U00013) (91 aa); Fasta score E(): 0, 98.9% identity in 91 aa overlap. Contains Pfam match to entry PF00121

4.6%; Score 23; DB 1; Length 312050;

Query Match



```

Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GCCGACGGCAGCTCCGCCGACGA 230
|||||
Db 112474 GCCGACGGCAGCTCCGCCGACGA 112452

RESULT 11
BD262946 7887 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION
DNA sequences for enzymatic synthesis of polyketide or
heteropolyketide compounds.
ACCESSION
BD262946
VERSION
BD262946.1 GI:33072714
KEYWORDS
JP 2002527067-A/11.
SOURCE
Polyangium cellulosum
ORGANISM
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 7887)
AUTHORS
Beyer, S., Bloeker, H., Brandt, P., Cino, P.M., Dougherty, B.A.,
Goldberg, S.L., Hofle, G., Mueller, R.J. and Reichenbach, H.
TITLE
DNA sequences for enzymatic synthesis of polyketide or
heteropolyketide compounds
JOURNAL
Patent: JP 2002527067-A 11 27-AUG-2002;
BRISTOL MYERS SQUIBB CO
COMMENT
OS Sorangium cellulosum
PN JP 2002527067-A/11
PD 27-AUG-2002
PF 11-OCT-1999 JP 2000576029
PR 09-OCT-1998 DE 198 46 493-2
PI STEFAN BEYER, HELMUT BLOEKER, PETRA BRANDT, PAUL M CINO, BRIAN A
PI DOUGHERTY,
PI STEVEN L GOLDBERG, GERHARD HOFLE, ROLF JOACHIM MUELLER, HANS PI
REICHENBACH
PC
C12N15/09, C12N1/21, C12N9/00, C12P1/04, C12P7/26// (C12N1/21, C12R1: PC
01),
PC (C12N1/21, C12R1:19), (C12N1/21, C12R1:465), C12N15/00 CC DNA
sequences for enzymatic synthesis of polyketide or CC
heteropolyketide
CC compounds
FH Key Location/Qualifiers
FT source 1..7887
FT /organism="Sorangium cellulosum".
FEATURES
source
1..7887
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/db_xref="taxon:56"
ORIGIN
Query Match 4.4%; Score 22; DB 6; Length 7887;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CGACCGGGCGGCCCAAGGGCGT 41
|||||
Db 1904 CGACCGGGCGGCCCAAGGGCGT 1925

RESULT 12
AF321122 9858 bp DNA linear BCT 29-OCT-2002
LOCUS
DEFINITION
Streptomyces antibioticus simocyclinone biosynthetic gene cluster,
partial sequence.
ACCESSION
AF321122
VERSION
AF321122.1 GI:11321496
KEYWORDS
Streptomyces antibioticus
SOURCE
Streptomyces antibioticus
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 9858)
Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M. and
Heide, L.
Cloning and analysis of the simocyclinone biosynthetic gene cluster
of Streptomyces antibioticus Tu 6040
Arch. Microbiol. 178 (2), 102-114 (2002)
22109063
PUBMED
12115055
REFERENCE
2 (bases 1 to 9858)
AUTHORS
Galm, U., Li, S.-M., Schimana, J., Fiedler, H.-P. and Heide, L.
Direct Submission
Submitted (13-NOV-2000) Pharmazeutische Biologie, Pharmazeutisches
Institut, Mikrobiologie/Biotechnologie, Universitaet Tuebingen, Auf
der Morgenstelle 8, Tuebingen 72076, Germany
Location/Qualifiers
1..9858
/organism="Streptomyces antibioticus"
/mol_type="genomic DNA"
/strain="Tu 6040"
/db_xref="taxon:1890"
100..810
/gene="simK"
100..810
/gene="simK"
/notes="putative reductase; similar to NovK of novobiocin
biosynthetic gene cluster; involved in aminocoumarin
biosynthesis"
/codon_start=1
/transl_table=11
/product="SimK"
/protein_id="AAG34182.1"
/db_xref="GI:11321497"
/transl_table="MAPVAIVAGVGTGTVVAVRLAEGDVAVLCADGAGNADTVH
RVAVGERALAVETVTDVAVQAALDRVCAELGDPGLVNSVGVACGAPGEASED
WNAVGSRGLGLFTVSRVTDLMIRQGRGRIVTIANLIGPEGGPHDHTVLGMAGF
TRTVALLRAFDTVNAVAPPPVGVGQAPAAFTADPEGSRCATQAAEAVLFLVGG
ASVTGQTQVGGDEAAA"
847..2406
/gene="simL"
847..2406
/gene="simL"
/notes="amide synthetase; similar to novobiocin acid
synthetase NovL; involved in aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="simL"
/protein_id="AAG34183.1"
/db_xref="GI:11321498"
/transl_table="MEGNEHYVQIQLNTLRADPSGVALVHRDTPVIAGDLADTSAA
EAMRGSGVGVSGVIGILTDENTPATIVRYAANLLGATVHLFGVNAANPSDLISAE
QGGIVAEALPAMVVMVDAANLERARATREVPSPVLSGLGELGHDVLDLTDSPAGFR
PDAARDGDTAVTTFSSGTRGPKGTAWSPRVKADMAASARRAKATAPATVTHGN
GFVADDVIVSGGVTVLLPGDFDETVLSVARYQVNRVLAAPOLYALADHPETRTDL
SSVRLDFTVTGVAASPERSVAEKVFGSLMVCQVGTSTNIIISLLAGEHTDAGLRATV
GRLEMLRVTRDPPQERVLPTGETGEVWNSPMRMDHNDPEQTARTVRDGMIRTV
DVHGLDAGYLHLHGLAGVITKNGIKVFPVAVERSLLDHPDVAEAAVFGVNSDRVE
RIHAVVVRREGAGAGPEDLRQHVSSHLSPNHAPADIELRSSLPLIGFGKPKDLRLAD
AMARREATHGE"
2435..5428
/gene="simH"
2435..5428
/gene="simH"
/notes="peptide synthetase-like protein; similar to NovH of
novobiocin biosynthetic gene cluster; involved in
aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="simH"
/protein_id="AAG34184.1"
/db_xref="GI:11321499"
/transl_table="MAMPSSGSVSVVEVEGLVDVSVFAAAVDVLLKRPAPTAGLSWR
EVDKSGPDQDSRRTGLDRLATEDGEAVVVGAPPRLTLVRLRADAFQURLASDRAVLDA

```



```
PSHRVILGELLECAAPAGAVGAPVEDRVAVTAETGEGRAERAWRIIRLAEVDPTIV
APAGAGTSQPERVESELPDILAEIRLLVRTDDRSLDTIYHAAWGVLVSLWNRN
VVFGLRSRDRFVAETAVGFGSEIVFVPCAVPAEPFADLLSKTELDGILIAHESA
LGAARIALVGLPRFDTAVGVEDAGDLPVPAPEGLPLGRVVAQWEDFGSALAVV
RRSGEGLRLNDFPELFTQADAEIRLAAAFRTTITADEENQVGRVQLDDHEQRR
ILASDADADGVAVTLVLELLOTVARTPGNTVVSFGQHI SYDELNRVNRRLARILA
DRGAPGOLVAVLPRSTIEMTAVLAVAKTGAGFLPIDPGVPADRIAYMLDAGPAPV
CTTRAAVILQDVSGIULDPAVAAEAGISDEDPVASEHAPVPPAALAVILYTSQ
STGPAGVAVTHAGLNUVAAKVERMDVDEQSRILOPASPFDAMFTELLATIAGAT
LVPPPGTLAGDHAEVLVAERITHVLVLPVVAASVSPESLPDLKRSULVIAEAGSDL
IARWAPGRVINYAGTEPTCATMSEPLSADATPIGCPITPGAACYVLDALRPVPA
GVPGELYGGAGLGRVPCMTAERFVNPFAAGDSRMVYTGDLVSWRSDGGLDFL
GRSDQVLCGRFIRLGEVERVLTNHPGVDRVAVVRDGGAGGRIIVVHLIPSAGAAP
TMAELREAGRLPDPMPVPGFVLDLDAFLPTENGKLDRLALPEPDPAAPSGRAPSTP
BEIALUSRIFSEVLGVAERVDADRNFPELGGNSMVALTVIQKARKAGLAISPKDLVWNP
VEALAAIAGRTDQAPRAEATTAVK"
gene
5440..6690
/gene="sim1"
CDS
5440..6690
/gene="sim1"
/note="cytochrome P450; similar to NovI of novobiocin
biosynthetic gene cluster; involved in aminocoumarin
biosynthesis"
/codon_start=1
/transl_table=11
/product="Sim1"
/protein_id="AAG34185.1"
/db_xref="GI:11321500"
/translation="MNPREMLSPDVLIEDLNDRLQHADYDLSSEWVRYLRARPPFYQ
TARGQPGFWVTRADCTAVYDKNTFTAERGNVLPVTLGAGDSASRTMLALTQDGR
HTQVNLKMAKSPKMLNIGOSLRTVVDGLLRDALEKGECDPFADVSKRSPVLAICD
LLAVQDEURYLSTAHALSADAEADATADNWTAKNETLLYFADLAEKRSSGHNVD
VSLATSSLEGPSLGGELMANYGLMIGDSTGHATYGLRALIHPDQWMLRNG
EADLQATBEVLRWTVPFHLGATADVNVNGKQIRAGEIVSVWFASANDREVEFR
DADFLNRTNKHITFAFGPHFLGHLRYARMEVEALDGLRRWDDIIOQTGPEKLIY
SSILOIGSFPALLKPDRRVPQT"
6727..6939
/gene="simy"
CDS
6727..6939
/gene="simy"
/note="MbH-like protein; similar to CouY of coumermycin
biosynthetic gene cluster; involved in aminocoumarin
biosynthesis"
/codon_start=1
/transl_table=11
/product="Simy"
/protein_id="AAG34186.1"
/db_xref="GI:11321501"
/translation="MANPFDQRGSLFLNRNAEEQVSLWPAFAGVPTGWQVAKGNR
AACLATEAWTLDRPKSLIDATDPS"
6950..7771
/gene="simj1"
CDS
6950..7771
/gene="simj1"
/note="putative 3-oxoacyl-[ACP]-reductase; similar to NovJ
of novobiocin biosynthetic gene cluster; involved in
aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="Simj1"
/protein_id="AAG34187.1"
/db_xref="GI:11321502"
/translation="MNPWPATRADRNICREGARMSLGDATVAIVTGAGRGICATA
QRLAAGATVAVDTRADTADTVASIRAAAGRALGICDVTVDLVEAAVDRVAFB
GRDLVNNAGVTRDSLVFMWDEDMPTVDVHLNGAARTVRAARQMWQGGRIYVN
LSSITAAALNOGOANYATAKAIQGYTRTLAVELPHGHTVNAIAPGFIATITDDTAR
RMGSDPVALRKAVASRVPMMRVGRPEDIAGLAVFLAGPDASYLTGQTIYVGGGPQ"
7941..8696
/gene="siml"
CDS
7941..8696
/gene="siml"
/note="putative response regulator; involved in
aminocoumarin biosynthesis"
```

```
/codon_start=1
/transl_table=11
/product="siml"
/protein_id="AAG34188.1"
/db_xref="GI:11321503"
/translation="MRBRPSKLITGGSSVVQLITVDDLOFEEELADLLPNEGFOARRE
DGIAGLTKCRSDVILVDLAISDTDFEVCRAIRAVSHVPIIIVSARDDDELQILSL
KLGADDITVPCGNQRLVARMEALIRRARNAWESGVPLASAEAGVQFGFVRIDLQR
RSVTRDGEVPLTRKEFDMLALLATDPGRVTRQIMPEVWGHGDAGTRTLGVHVTG
IRKLRMLPELVETVGVGFRLANHTLSLSDQAAA"
9018..9758
/gene="simj2"
CDS
9018..9758
/gene="simj2"
/note="putative 3-oxoacyl-[ACP]-reductase; similar to NovJ
of novobiocin biosynthetic gene cluster; involved in
aminocoumarin biosynthesis"
/codon_start=1
/transl_table=11
/product="simj2"
/protein_id="AAG34189.1"
/db_xref="GI:11321504"
/translation="NVHNRGVALTGGSRGIGRAVILRLVREGFDVAFCYOREAVAA
ADLEKAVAEELGSRVARRVDSDAVQVKEVRDVESELGPVEVVTSGIVRDGDLVM
MODKMDQVIDVNLNGTNNVCRAAVQGMKRRSGCLINLSSVAGVSGHATQSNYSASK
AGIIIGTKSLSKELGRYGRANVAPGFIITDMVAGISDGLFETKAVAGVALGRFGRPD
EVADAVAYLAGAEYVTGAVLQIDGGLSI"
ORIGIN
Query Match 4.4%; Score 22; DB 1; Length 9858;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 CGACCGGGCGGCCCAAGGGCGT 41
DB 4134 CGACCGGGCGGCCCAAGGGCGT 4155
RESULT 13
AX707120
LOCUS AX707120 9858 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 6 from Patent WO03014352.
ACCESSION AX707120
VERSION AX707120.1 GI:29563425
KEYWORDS Streptomycetes antibioticus
SOURCE Streptomycetes antibioticus
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Heide, L. and Li, S. M.
AUTHORS Nucleic acids for aminocoumarin biosynthesis
TITLE Patent: WO 03014352-A 6 20-FEB-2003;
JOURNAL Universitaet Tuebingen (DE)
FEATURES
source
1..9858
/organism="Streptomyces antibioticus"
/mol_type="unassigned DNA"
/db_xref="taxon:1890"
ORIGIN
Query Match 4.4%; Score 22; DB 6; Length 9858;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 CGACCGGGCGGCCCAAGGGCGT 41
DB 4134 CGACCGGGCGGCCCAAGGGCGT 4155
RESULT 14
AX803758
LOCUS AX803758 18876 bp DNA linear PAT 25-NOV-2003
```

DEFINITION Sequence 42 from Patent WO03060127.  
ACCESSION AX803758  
VERSION AX803758.1 GI:39520925  
KEYWORDS Streptomyces refuineus  
SOURCE Streptomyces refuineus  
ORGANISM Streptomyces refuineus  
Bacteria; Actinobacteria; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 Farnet,C.M., Staffa,A. and Zazopoulos,E.  
AUTHORS Genes and proteins involved in the biosynthesis of lipopeptides  
TITLE Patent: WO 03060127-A 42 24-JUL-2003;  
JOURNAL Ecopia Biosciences Inc. (CA)  
FEATURES Location/Qualifiers  
source 1..18876  
/organism="Streptomyces refuineus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:23296"  
ORIGIN  
Query Match 4.4%; Score 22; DB 6; Length 18876;  
Best Local Similarity 100.0%; Pred.No. 9.2; 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0;  
QY 20 CGACCGGGCGGCCAAGGCGT 41  
|||||  
Db 13034 CGACCGGGCGGCCAAGGCGT 13055  
RESULT 15  
AF324838 34869 bp DNA linear BCT 14-NOV-2003  
LOCUS Streptomyces antibioticus simocyclinone biosynthetic gene cluster,  
DEFINITION partial sequence.  
ACCESSION AF324838  
VERSION AF324838.1 GI:12744817  
KEYWORDS Streptomyces antibioticus  
SOURCE Streptomyces antibioticus  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 34869)  
AUTHORS Trefzer,A., Pelzer,S., Schimana,J., Stockert,S., Bihlmaier,C.,  
Fiedler,H.P., Welzel,K., Vente,A. and Bechtold,A.  
TITLE Biosynthetic gene cluster of simocyclinone, a natural multihybrid  
antibiotic  
JOURNAL Antimicrob. Agents Chemother. 46 (5), 1174-1182 (2002)  
MEDLINE 21955989  
PUBMED 11959542  
REFERENCE 2 (bases 1 to 34869)  
AUTHORS Trefzer,A. and Bechtold,A.  
TITLE The simocyclinone biosynthetic gene cluster isolated from  
Streptomyces antibioticus Tue6040  
Unpublished  
JOURNAL 3 (bases 1 to 34869)  
REFERENCE Trefzer,A. and Bechtold,A.  
AUTHORS Direct Submission  
TITLE Submitted (29-NOV-2000) Pharmazeutische Biologie,  
JOURNAL Christian-Albrechts-Universitaet Kiel, Gutenbergstr. 78, Kiel  
24118, Germany  
FEATURES Location/Qualifiers  
source 1..34869  
/organism="Streptomyces antibioticus"  
/mol\_type="genomic DNA"  
/strain="Tue6040"  
/db\_xref="taxon:1890"  
/clone="cosmid 5J10"  
1337..2812  
/gene="simA7"  
1337..2812  
/gene="simA7"  
/note="involved in polyketide formation"  
/codon\_start=1

/transl\_table=11  
/product="putative oxygenase SimA7"  
/protein\_id="AAK06782.1"  
/db\_xref="GI:12744818"  
/translation="MDAQIVVVGAGPAGLMLAGELRLAGVDVVVLERLAQPTGESRGL  
GFTARTLEVFDORGLPRLGDIETVLDFGLGGLDPLDFILGPAHSGGVKGVPAVTEA  
MLDEWALELGADIRGHELIJSTONGDFVEVEVATSDGPORLRTAPLVGCDGRTVTR  
KAAGDFPGTSATMEMFLGVRGCDITPRMIGETVPGMWAAPIGNGVDRIIVCEHG  
TPARRRDEPFGWEEAAWQRLTGODISAGEFPWLSAFGNPARQVTERRGRVLLAGD  
AAHTLPAGGQGMVTSIQDVNVNLGWKLAATVKGTGAPEGLLDSHYTERHPVGRRLMNT  
OAGLFLSGAEMOPLDLHMGSELIITFENVRHLAQMSVGLSIRYDLGPDHPLGLGRI  
PHLELVGAGKTTSTELLHPARGVLLDLADDAGLRETAAPWADRPADRVVTVTVHGAADA  
PQYEPADALLIRPDGHVAAAGPVALSEALARSFGAPSR"  
2862..3188  
/gene="simA4"  
2862..3188  
/gene="simA4"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative cyclase SimA4"  
/protein\_id="AAK06783.1"  
/db\_xref="GI:12744819"  
/translation="MHSTLIIVARMSPRDDQDVARLFSEPDRTMPHLMGTRRRQLFSY  
RGLYFHLQDFDDEHGEALEAKSHPSFVRISEDLPFIQAYDPDPTWRSADAMARRF  
YOWSAQ"  
3185..4453  
/gene="simA1"  
3185..4453  
/gene="simA1"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative ketosynthase SimA1"  
/protein\_id="AAK06784.1"  
/db\_xref="GI:12744820"  
/translation="MRRRVITGVGVMAPEGGLGVKNFWSILLSEGRATRGITFFDPAP  
FRSRVAEIDFFPEKHGLGPKIRMDRAAQLAVIATREAVADSGLLDLDTPDRSGV  
TVGSVAETGTTQDEYRVVSDGLRGLVDHRYAPPHLYGHFVFPFAAEVAMASGSEG  
PTVSTVGTCTGLDGLHVELIREGSADVMTAGATDAPISPTMACFADAKATPRN  
DIPAHASRPDSRNGFVLGEHVALVLEELGAAQRGARIYAEIIGYASRCNAFMT  
GLRPGREMAEIRVALDEARLNPEDIDVYNAHSGTQKNDHRETAFAKRSILGDHAYR  
TPVSTIKSMVSGSLGAISIEIAACVLMAESVPTANLHDPPEPCDLDYVITARE  
ARLDRLVSGSGFGFQSAWLVTRPERNTA"  
4450..5664  
/gene="simA2"  
4450..5664  
/gene="simA2"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative ketosynthase SimA2"  
/protein\_id="AAK06785.1"  
/db\_xref="GI:12744821"  
/translation="MTTSVVTVGVGVAAPNGLCTADYWAATREGRSGIGRVTFRDPQ  
YPSRLAGEVPGFAEDHRSRLPQTDHTRLALVSDWALQDAGIRPEELPEVAGV  
VTASSAGGFEFGQNELRALMSKGSYVSAFVSWGQALTSGRMSTDDDDPATATY  
SDQAGDLAQARQIRRGSGVIVSGGIDASICSMWGQALTSGRMSTDDDDPATATY  
PFGAAGHVPGEGGALLVLEEAAGAAARGVYRIAGYGTTFDPKPGSGREPLERR  
AIELATDAGLAPADIDVVFADAAVFDLREARLVEARLVFVGPRGLPVTAPKSGMTGRL  
NSGAGSLDVATALLARDGVIPTNTVTAODDYELDLVTAQARSARLSALVWARGHG  
GFNSALVVDAA"  
5713..5979  
/gene="simA3"  
5713..5979  
/gene="simA3"  
/note="involved in polyketide formation"  
/codon\_start=1  
/transl\_table=11  
/product="putative acyl carrier protein SimA3"  
/protein\_id="AAK06786.1"  
/db\_xref="GI:12744822"  
/translation="MQTTEFLDLRLRDGAGSEDVDLEGDIIIDVAFEDLGYESI

```
gene
CDS
/ALLETCSRIERYGISLDDDAVSEAPTALVAIVNTHLIAVAG"
5999..6784
/gene="simA6"
5999..6784
/gene="simA6"
/notes="involved in polyketide formation"
/codon_start=1
/transl_table=11
/product="putative ketoreductase SimA6"
/protein_id="AAK06787.1"
/db_xref="GI:12744823"
/translation="MPGQKQVALVTGATSGIGLAVVRALAAQHQHVFCCARNAPKVA
VTIKELRGDLDVDCSDVRAADVKAFVAAAVRFGPVDVLVNNAGRSGGVTFAEL
TDELWADVIENLTSTVPMVTRVLTAGMREKSGRIVNIATAGQGVILCAPYSAS
KHGVYFTKAGNELAPTGITVNAVCPCGVETPMARVROGVAAAVGTSEBAILEKFPQ
AKIPUGRSTPEVAGLVGLVSDTAAISITAQALNVCGGLGNF"
6810..7757
/gene="simA5"
6810..7757
/gene="simA5"
/notes="involved in polyketide formation"
/codon_start=1
/transl_table=11
/product="putative cyclase/dehydrase SimA5"
/protein_id="AAK06788.1"
/db_xref="GI:12744824"
/translation="MSDSGISHVHTIEVDAPATDVYRLIAEVENWPRIFGPTIHVDH
VERGARTIRIWTANGAAKSWTSRRELDPEALRIDFQFVSAPPVAMGAWLIEP
VSEGCARVRLHDYRAIDDAPELDIAEWQVDRNSRSELAALKANREVVVGADELML
SFQDSVRIDGSAKDYDFNEADRWTRPHEAVELTEQTFGLQVLMETRAKDGST
HTSGVRVCFPHRIVYKQTVLPALMTLHTGYLLEEDGEGTVATVTSQHTVIINTAIR
ATLGSAGVAERQVYQTALSTNSRATLGHAKDYAERRD"
7763..9247
/gene="simA8"
7763..9247
/gene="simA8"
/notes="involved in polyketide formation"
/codon_start=1
/transl_table=11
/product="putative oxygenase SimA8"
/protein_id="AAK06789.1"
/db_xref="GI:12744825"
/translation="MTGPQTDVILVIGDPDPLMLAGELLGGARVVVAEBLITPTESR
ARCRPTWEILDGLTGLRGVGDGTHFGGILRLDGOOSPVAGLWKVQPSRVE
EILGAWAIGLGRVREGRBLTGLREDGAVVEAELSCDSTRELLRRAVYVDCDGNST
VRLAGFEPAGSARRELRADVAGIDVNRPERLPGLATAARPDCGITRUMVHEF
GRTATPRAEPDPEDEWARTQWRVTGSDISHGTFVWINSFGDASRQVTEFRGILLAG
DAHAQMVGQGLNLQDAFNLGWLKAAEICGWAPPGLLDSYHEERHAGVGRVLSD
IAAQALLLGGAEVRLSVYGRLLGLDAVRVPFGRHRRHLSSTPTTRVHPLIGA
RLPDAALDTEAGATAWALLRAGTGLILDVAPAPGHVDILRAVADGWADRVRYVAARP
RAGAVAPTCAVLVRPDGHVVHWMADGGADRLRAALVQWFGSADRT"
9326..10090
/gene="simA9"
9326..10090
/gene="simA9"
/notes="involved in polyketide formation"
/codon_start=1
/transl_table=11
/product="putative reductase SimA9"
/protein_id="AAK06790.1"
/db_xref="GI:12744826"
/translation="MKRLAGKTALVTGSSRGIGRAVALRLAREGATVAVHYADNADAA
KATLAEBSGGAFLVRLRGLVPGVDVTLFQGLTAGLREATGKEELINLVNAGVNG
GVXPEDLTPEIFQOLIAVNAKAPLFLTORVLSIMPDGRLINIGTGLLRVANDEVAY
SMKSGALDQISLHFAFLRLGNITVNTVSPGTTNNGSPIDFPEAREFMANMAFKRI
AEPEDVDGVVAFLASDDARWITGTFVDASGGLTLLG"
10221..10793
/gene="simA10"
10221..10793
/gene="simA10"
/notes="involved in polyketide formation"
/codon_start=1
/transl_table=11
```

```

/product="putative reductase SimA10"
/protein_id="AAK06791.1"
/db_xref="GI:12744827"
/translation="WSSARLR LAVLGSITREGRLCPTVANWFAGQGRHQGFDDLDLVD
PYGLDIPVAPSPPEPAOVVEIGORLAADAFVITPEYNHSFSAPLKSLIDWYVQWR
AKPVFVSYGGLAGLRAVEQLRLVFAELHAATVDRDITVSFNAAQLFDEAGDPRPEG
CNTAAKTMLDQLHWWGRALRQARLERPYQG"
10801..11466
/gene="simA11"
10801..11466
/gene="simA11"
/notes="involved in polyketide formation"
Query Match 4.4%; Score 22; DB 1; Length 34869;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 CGACCGGCGGCCCAAGGCGT 41
DB 25206 CGACCGGCGGCCCAAGGCGT 25227
RESULT 16
BD262936 37856 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION
DNA sequences for enzymatic synthesis of polyketide or
heteropolyketide compounds.
ACCESSION
BD262936
VERSION
BD262936.1 GI:33072704
KEYWORDS
JP 2002527067-A/1.
SOURCE
Polyangium cellulosum
ORGANISM
Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Sorangineae; Polyangiaceae; Polyangium.
1 (bases 1 to 37856)
Beyer, S., Bloeker, H., Brandt, P., Cino, P.M., Dougherty, B.A.,
Goldberg, S.L., Hofle, G., Mueller, R.J. and Reichenbach, H.
DNA sequences for enzymatic synthesis of polyketide or
heteropolyketide compounds
Patent: JP 2002527067-A 1 27-AUG-2002;
BRISTOL MYERS SQUIBB CO
OS Sorangium cellulosum
PN JP 2002527067-A/1
PD 27-AUG-2002
PF 11-OCT-1999 JP 2000576029
PR 09-OCT-1998 DE 198 46 493.2
PI STEFAN BEYER, HELMUT BLOECKER, PETRA BRANDT, PAUL M CINO, BRIAN A
PI DOUGHERTY,
PI STEVEN L GOLDBERG, GERHARD HOFLE, ROLF JOACHIM MUELLER, HANS PI
REICHENBACH
PC
C12N15/09, C12N1/21, C12N9/00, C12P1/04, C12P7/26// (C12N1/21, C12R1: PC
01), (C12N1/21, C12R1:19), (C12N1/21, C12R1:465), C12N15/00 CC DNA
sequences for enzymatic synthesis of polyketide or CC
heteropolyketide
CC compounds
FH Key Location/Qualifiers
FT source 1. 37856
FT /organism="Sorangium cellulosum".
FEATURES
source
Location/Qualifiers
1. 37856
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/db_xref="taxon:56"
ORIGIN
Query Match 4.4%; Score 22; DB 6; Length 37856;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 CGACCGGCGGCCCAAGGCGT 41
|||||
```



```
/transl_table=11
/product="sim7"
/protein_id="AAL15585.1"
/db_xref="GI:16224004"
/translational="MTGPQDVLVIAGPVGLMLAGELLGGARVVVABRLTEPTTES
RASTLHARWEILDQRLTGLRGEVPGDGTGHFGGIPRLRDGQSQPYAGLWKPQSRV
EELGAWAIGLAEVREGLTGLRGEVAVAEALSCEGTRELLRARVYVCGDGENS
TVRLAGFEFAGSARELIRAVAGIDVPNRRFERLPGLAIAARRPPDITRMVYHE
FGTATPROAEPDPDMATWQRTVEDISHGTPVMNSFGDASQVTVYRGRILLA
GDRAHQMPVGGQALNLQDANLGNKLAIEICGNAPGGLDSYHEERHAGVSRLS
DIAAQLLLGGPEVLRSVVGRLLGLDAVRDLHATGVGGIGIQIYSHHAGDPLIGA
RLPDALTEAGATATWALRLRAGTGILLDVAPEAPGHVDLRAVADGWDRVRYAARP
RAGAVAPTTCGAVLVRPDGHVWADGGADRRLRAALVQWFGSADRT"
5041..5805
/gene="sim8"
CDS
5041..5805
/notes="oxidoreductase"
/notes="sim8"
/codon_start=1
/transl_table=11
/product="sim8"
/protein_id="AAL15586.1"
/db_xref="GI:16224005"
/translational="MKRLAGKTALVTGSSRGIGRATALRLAREGATVAVHADNAAA
KATLAATEERGGRAFLVRLAGTVPDVTTLFQGLTAGLREATGKEELNIIIVNAGYMG
GVKPEDLTPEIFORLIANAKAPLELTORVLSIMPQGRILINTGTGLLRVANPDVAY
SMKSGALDQISLHARLLGPRNLTWNTVSPGINTNGSPIDFVPEAREFMANNAFXEI
AEPEDVDVAVFLASDARWITGTFYDASGTLIG"
5936..6508
/gene="sim9"
CDS
5936..6508
/notes="reductase"
/notes="sim9"
/codon_start=1
/transl_table=11
/product="sim9"
/protein_id="AAL15587.1"
/db_xref="GI:16224006"
/translational="MSSARLRLAVLGTREGRLCPTVANWAGQGHGGFDLLVD
PYGLDIPVASPPPAQVVIQORLAADAFVIVTEYNHSFSAPLKSLIDWYVQWR
AKPVGVYSGLLGGRAVQLRLVFAELHAATVDTVSPHNAQLFDRAGDRDPBG
CNTAAKTMLDGLHWGRLARLERPYQG"
6516..7202
/gene="sim10"
CDS
6516..7202
/notes="putative phosphopantetheinyl transferase"
/codon_start=1
/transl_table=11
/product="sim10"
/protein_id="AAL15588.1"
/db_xref="GI:16224007"
/translational="MATESVQPAEAVOLWVRPEVPRDELALDSEERRRAEAFRG
SADRMVLSAHLARLVLAARLGHVAPQVEVIRVDRNGRPTLPGRDPPFFHFSLSHAG
LALLGTAPVIGVDVQTLTKRTADLCRLHAEQELASVQPSARAAHFTLWTRK
EAYLKLGVGLSRLAADYLGVRPDTFAGTWVLDVPCGSHLASAAVQSRSLPVTLL
RWLLRSADHV"
7195..8781
/gene="sim11"
CDS
7195..8781
/notes="carboxyl transferase/decarboxylase"
/codon_start=1
/transl_table=11
/product="sim11"
/protein_id="AAL15589.1"
/db_xref="GI:16224008"
/translational="MSDVPSIPADPAPTTCRLRELRALROAVADGPDPRATEAQAHK
GKLTARRVELLDGSGFHEVEPLRRHRAQGLFLEARKPHDTGTVVCGTWGTVHGTVFV
YKLTARRIFGGLGAHQKTHKINDMALAGAPLVSLNDGAGARLQGVYALAGYGGI
FORNTRASGIPQISLWMLGPCAGGAAYSPALDTFVFMVRETSQMFITGPDVQAVTGE
KISQNGLGADVAGSVGSHFVYDQSCIEEYVYLLSLPNSNREMPVQLSDDDFV
DRRNDSLDLVPADGNRPYDMRKVIEEVVDYGEYLEVHERWATNVLCVLRIDGHVTC
```

## gene

```
8796..9014
/gene="sim12"
CDS
8796..9014
/gene="sim12"
/notes="hypothetical 7.6 kDa protein"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="sim12"
/protein_id="AAL15590.1"
/db_xref="GI:16224009"
/translational="WTSDAWTDAIRVESGHLEPEBELAALTVLLAIRGAPDRAAGD
GGSLVAATWRRUERANGFQSPSWRTAC"
9056..9829
```

## gene

```
Query Match 4.4%; Score 22; DB 1; Length 39428;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 20 CGACCGGGCGGCCCAAGGCGGT 41

Db 20922 CGACCGGGCGGCCCAAGGCGGT 20943

## RESULT 19

```
AX803750
LOCUS 61944 bp DNA linear PAT 25-NOV-2003
DEFINITION Sequence 34 from Patent WO03060127.
```

ACCESSION AX803750

VERSION AX803750.1 GI:38520921

KEYWORDS Streptomyces refuineus

SOURCE Streptomyces refuineus

ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE 1 Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS Farnet, C.M., Staffa, A. and Zazopoulos, E.

TITLE Genes and proteins involved in the biosynthesis of lipopeptides

JOURNAL Patent: WO 03060127-A 34 24-JUL-2003;

Ecopia Biosciences Inc. (CA)

FEATURES Location/Qualifiers

source 1..61944

/organism="Streptomyces refuineus"

/mol\_type="unassigned DNA"

/db\_xref="taxon:223296"

## ORIGIN

```
Query Match 4.4%; Score 22; DB 6; Length 61944;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 20 CGACCGGGCGGCCCAAGGCGGT 41

Db 18094 CGACCGGGCGGCCCAAGGCGGT 18115

## RESULT 20

```
BX664737
LOCUS 144560 bp DNA linear HTG 06-NOV-2003
DEFINITION Homo sapiens chromosome X clone CTD-2511C7, WORKING DRAFT SEQUENCE,
4 unordered pieces.
```

ACCESSION BX664737

VERSION BX664737.5 GI:38201324

KEYWORDS HTG; HTGS PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 144560)

AUTHORS Heath,P.  
TITLE Direct Submission  
JOURNAL Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Nov 7, 2003 this sequence version replaced gi:38091229.  
COMMENT ----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BK2511C7  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 143823 bases at least Q40  
Consensus quality: 144044 bases at least Q30  
Consensus quality: 144181 bases at least Q20  
Insert size: 144260; sum-of-contigs  
Insert size: 165629; 9.1% error; agarose-fp  
Quality coverage: 18.13x in Q20 bases; sum-of-contigs Quality  
Coverage: 16.87x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 48654: contig of 48654 bp in length  
\* 48555 48754: gap of 100 bp  
\* 48755 139049: contig of 90295 bp in length  
\* 139050 139149: gap of 100 bp  
\* 139150 142135: contig of 2986 bp in length  
\* 142136 142235: gap of 100 bp  
\* 142236 144560: contig of 2325 bp in length.  
FEATURES  
source Location/Qualifiers  
1..144560  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone\_lib="CIT-HSP-D2"  
/clone\_1ib="CIT-HSP-D2"  
misc\_feature 1..48654  
/note="assembly fragment:01776  
fragment\_chain:1"  
misc\_feature 48755..139049  
/note="assembly fragment:04387  
fragment\_chain:1"  
misc\_feature 139150..142135  
/note="assembly fragment:04669  
fragment\_chain:2"  
misc\_feature 142236..144560  
/note="assembly fragment:04697  
fragment\_chain:2"  
ORIGIN  
Query Match 4.2%; Score 21; DB 2; Length 144560;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 62 GCCTCTGCTCGTGGCTTCT 82  
|||||  
Db 56213 GCCTCTGCTCGTGGCTTCT 56233  
|||||  
RESULT 21  
AL646066  
LOCUS 189050 bp DNA linear BCT 11-APR-2003

DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;  
segment 10/19.  
ACCESSION AL646066 AL646052  
VERSION AL646066.1 GI:17428714  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE 1  
AUTHORS Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
Chandler M., Choienne N., Claudel-Renard C., Cunnac S., Demange N.,  
Gaspin C., Lavie M., Moisan A., Robert C., Saurin M., Schiex T.,  
Signier P., Thebault P., Whalen M., Wincker P., Levy M.,  
Weissenbach J. and Boucher C.A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
JOURNAL Nature 415 (6871), 497-502 (2002)  
MEDLINE 21681879  
PUBMED 11823852  
REFERENCE 2 (bases 1 to 189050)  
AUTHORS Boucher C.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex. Laboratoire de Genétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
Location/Qualifiers  
1..189050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GMI1000"  
/db\_xref="taxon:305"  
170..3325  
/gene="RSC1698"  
/note="synonym: RS02885"  
170..3325  
/gene="RSC1698"  
/function="elements of external origin; phage-related  
functions and prophages"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
Predicted by Codon usage  
Predicted by Homology  
Predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transal\_table=11  
/product="PROBABLE BACTERIOPHAGE PROTEIN"  
/protein\_id="CAD15400.1"  
/db\_xref="GI:17428715"  
/db\_xref="SPTREMBL:O8XYR0"  
/translation="MRNIGYGGKGGSSSTTEAPDLSHVAARVLVLSEGEIA  
GLVNGLSIYELGTPLANADGTLNFQNVAVDYPGQDQQAIFGPFSESESAVGVEL  
TAMPVRAIANTQLSAVRVQLSRADTSNGINGYRVEYADLVADGALQOV  
LATAFGKTKSVTRTRHLELPAKGTWIRVRIIPNANGSIADVTRLESTAEVID  
AKLRYPNSALIGRIDARQESPVSPTSYHLRGLIIRVPSNYDPATRYTGLMGDTFKV  
AYSNFAPVFIIVLHRCGLGDRKADMDKWLQYIGYCDLVDVDDGGQEPRT  
CNCYLQORNDAYVLQDLASVFRGMAWASNVAVADMPASAYLFHPGNVDGRT  
YAGSARRAKTVALVNDPDQYRAKVEHVEDGDGIARYSIQQTETAFCTGSQA  
VRVGOMILLTSLRETIVTSFKGLDAVMPGAIIEVADPARAGTNGRVRGADGT  
VVVDPRVTVREGDTLLNLPDGTAAQRTISRVDSQSLTVSADMSQPVQAQAVSVESA  
DLATQLFRVSVTESDGLTFDISALEYNSQSKFSAIDNGTRIEARPSVLPPSVQAPPT  
DVTLATYVVVDQGIATITMMVAQQAASAVATVEWRDNGEWTRAGTRGQSLVERN  
IYAGTYVARVRAINLADVASLPAISAEITLTGKTSPPPVVGTLLIATAIVFGIRLDWAF

```

PTGLDVBTELMWSKAPDRASATKGLGDFAPFANTHTLMGLAAGATFFFWARLVKSG
NVGAWPGSNGVLGSSDRADILSYMGQICKTQLAGDLSAVDSIDPPWAGSDEDF
AGDGTVAGIVSTQSVLEARNRAIAQVTTMOATVAQSAAVQAQQAQVAEQGRLAA
MYTKTQIAANGRIYLAGIFGVGNKGVISQVLLIAADRPAVLHPNGNNVFTFPVVO
GQVPMDSATADGTITTSKIGDSIQSNFVAGTGWRLSKSLFENNNGAGGRRVD
TSAUQIYDANGTLRLVGVW"
/ gene .5188
/ gene="RSC1699"
/ note="synonym: RS02888"
/ gene="RSC1699"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD15401.1"
/ db_xref="GI:17428716"
/ db_xref="SPTREMBL:Q8XYQ8"
/ translation="MPAGLVYGERGVQIIEAYGNLALRQKSTVTPGSGRGSTIFS
ASRPVCIQSANFVALIGSSNGADWTVKFSAPAQAPFTVYVDFEAPSPGQGLQVF
KADGTLAFLDGLAYLKVGSVTPPGSAPAYQAQWASPLTAGNYAACLSYTRGTIYAV
PGDTLFLVADHYVTTANGAGLKLOYTHRGANNQDGSGLKIDTANPPQIILVDVSVL
"
/ gene .4878
/ gene="RSC1700"
/ note="synonym: RS02889"
/ gene="RSC1700"
/ EC_number="3.2.1.-"
/ functions="elements of external origin; phage-related
functions and prophages"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
/ protein_id="CAD15402.1"
/ db_xref="GI:17428717"
/ db_xref="GOA:Q8XYQ8"
/ db_xref="SPTREMBL:Q8XYQ8"
/ translation="MPYTPSVLGGNNLAFLDMVGFSEGTNGRQPTDRDGYDVLVG
GGLFVSADHPRLVLPRLGKSTAGRYQLLARYDPYRRLQLRLVDFGPAQDAIA
VQOIRREGALADIQAGRLAAATKCNWASLPGAGYGGQEQKFEALRAHYLLCGQEQ
VKG"
/ gene .5186
/ gene="RSC1701"
/ note="synonym: RS02890"
/ gene="RSC1701"
/ function="miscellaneous; unknown"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
/ protein_id="CAD15403.1"
/ db_xref="GI:17428718"
/ db_xref="SPTREMBL:Q8XYQ7"
/ translation="MTESMAVAKAVKVGSAALGSVIALRFLPCTWORMLSFVSSLGI
GCLAGGAVERFLLVGSGSYTHMLAVASAAVFLGLAVNNNAQIPEILTLDRRFLAKE
"

```

```

5189. .5464
/ gene="RSC1702"
/ note="synonym: RS02891"
5189. .5464
/ gene="RSC1702"
/ function="miscellaneous; unknown"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE TRANSMEMBRANE PROTEIN"
/ protein_id="CAD15404.1"
/ db_xref="GI:17428719"
/ db_xref="SPTREMBL:Q8XYQ6"
/ translation="MLTSINQVANALIFVGLWAVLTHQVPTTGGLVLAIVNFAA
LGNIVAHGTCHSWPEVALNVALCVWVAFWRVRFRFSALGKGRG"
/ gene="RSC1703"
/ note="synonym: RS02892"
5464. .6009
/ gene="RSC1703"
/ EC_number="3.4.-.-"
/ functions="elements of external origin; phage-related
functions and prophages"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE
PROTEIN"
/ protein_id="CAD15405.1"
/ db_xref="GI:17428720"
/ db_xref="GOA:Q8XYQ5"
/ db_xref="SPTREMBL:Q8XYQ5"
/ translation="MPGKGTATAAAILGALVGAAGATWQAHRYDKDIATLQRD
QAEKDRQARADEAQLROERKAGLQGLGELDARHYEELNNKATERARLAAALAG
TQLSIRARCTAPADGVSTTSAGLGGGATALLYGEDAADIQALVADADAIAAL
RFFQGRERAIQTRDTSVDARR"
/ note="predicted by Homology"
/ complement(6117. .7304)
/ insertion_seq="ISRS018 or RS05932"
/ complement(6273. .7238)
/ gene="ISRS018"
/ note="synonyms: RS02893, RSC1704"
/ complement(6273. .7238)
/ gene="ISRS018"
/ function="elements of external origin; transposon-related"
Query Match 4.2%; Score 21; DB 1; Length 189050;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 CTACACGTCGGCGACGACCGG 26
Db 57379 CTACACGTCGGCGACGACCGG 57399
RESULT 22
HUMFLNG6PD/c HUMFLNG6PD 219447 bp DNA linear PRI 17-MAY-1999
LOCUS Homo sapiens chromosome X region from filament (FLN) gene to
DEFINITION Glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds.
ACCESSION L44140
VERSION L44140.1 GI:1203968
KEYWORDS 1A gene; 2.19 gene; ABP-280 gene; DNLI1 gene; DNase I-like protein
gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI
gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene;

```

XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin; emery-dreifuss syndrome; filamin; glucose-6-phosphate dehydrogenase.

## SOURCE

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 219447)  
AUTHORS Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N., Zuo,L., Heiner,C., Burrough,F.W., Ripetto,M., Schlessinger,D. and D'Urso,M.

TITLE Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCF/GCP and

G6PD loci Hum. Mol. Genet. 5 (5), 659-668 (1996)

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Submitted by: Ellison Chen,  
Advanced Center for Genetic Technology,  
Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln Center Drive, CA 94404 USA  
and  
David Schlessinger,  
Department of Molecular Microbiology and Center for Genetics in Medicine  
Washington University School of Medicine,  
St. Louis MO 63110 USA  
e-mail: ellison@genseq.apliddio.com and david@genetics.wustl.edu  
Note: Gene predictions were accomplished with runs of Grail  
versions  
1.1 and 1.2, coupled with fasta and blastx comparisons to genbank  
&  
non-redundant peptide libraries. Repeat analysis was accomplished  
via  
censor.

## FEATURES

## source

## Location/Qualifiers

1. .219447  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="X"  
4. .223  
/evidence=experimental  
/rpt\_family="Alu-Sp or Alu-Sq"  
226. .503  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
641. .931  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
1023. .1313  
/evidence=experimental  
/rpt\_family="Alu-Sx"  
complement(1446. .1536)  
/evidence=experimental  
/rpt\_family="Alu-J"  
3095. .3426  
/evidence=experimental  
/rpt\_family="CpG Island"  
complement(3663. .3952)  
/evidence=experimental  
/rpt\_family="Alu-Sb0 or Alu-Sb1"  
4912. .5046  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
5047. .5333  
/evidence=experimental  
/rpt\_family="Alu-Sb0 or Alu-Sb1"  
5346. .5628  
/evidence=experimental

repeat\_unit /rpt\_family="Alu-J or Alu-S"  
5632. .5805  
/evidence=experimental  
/rpt\_family="Alu-Sq"  
6363. .17698  
/evidence=experimental  
/rpt\_family="11Kb repeat1"  
6369. .6633  
/evidence=experimental  
/rpt\_family="Alu-Sx"  
complement(7121. .7409)  
/evidence=experimental  
/rpt\_family="Alu-Sp"  
complement(7358. .7432)  
/evidence=experimental  
/rpt\_family="Alu-Sb0 or Alu-Sb1"  
complement(7426. .7701)  
/evidence=experimental  
/rpt\_family="Alu-J"  
8306. .8592  
/evidence=experimental  
/rpt\_family="Alu-Sx"  
complement(8863. .9029)  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
complement(9031. .9317)  
/evidence=experimental  
/rpt\_family="Alu-Sq"  
complement(9318. .9442)  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
9486. .9773  
/evidence=experimental  
/rpt\_family="Alu-J"  
9996. .10132  
/evidence=experimental  
/rpt\_family="Alu-Sp"  
10159. .10255  
/evidence=experimental  
/rpt\_family="Alu-Sq"  
10271. .10367  
/evidence=experimental  
/rpt\_family="Alu-J"  
complement(10455. .10739)  
/evidence=experimental  
/rpt\_family="Alu-Sx"  
complement(10740. .10860)  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
complement(10866. .11156)  
/evidence=experimental  
/rpt\_family="Alu-Sx"  
11223. .11501  
/evidence=experimental  
/rpt\_family="Alu-J"  
11638. .13564  
/evidence=experimental  
/rpt\_family="CpG Island"  
complement(11884. .12163)  
/evidence=experimental  
/rpt\_family="Alu-Sb0 or Alu-Sb1"  
complement(12195. .12482)  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
complement(12668. .12945)  
/evidence=experimental  
/rpt\_family="Alu-J or Alu-S"  
complement(18360. .18648)  
/evidence=experimental  
/rpt\_family="Alu-Sb0"  
complement(18984. .45006)  
/gene="FLN"  
complement(18984. .19488)  
exon



```
/gene="FLN"
/number=48
/evidence=experimental
complement(join(19301..19488,19814..20017,20101..20319,
20483..20659,21361..21493,22033..22148,22336..22473,
22633..22899,23005..23127,23224..23376,23453..23656,
23748..23909,24006..24179,24367..24495,24603..24743,
24833..24935,25068..25163,25277..25524,27701..27724,
27884..28073,28649..28805,28895..29018,29434..29604,
29696..29856,29934..30096,30182..30355,30440..31037,
31758..32030,32120..32237,32429..32598,32692..32782,
32868..33028,33111..33234,34472..34615,34709..34822,
34976..35169,35271..35407,35586..35709,35798..35935,
36473..36673,36757..36919,37011..37088,37181..37299,
37846..37993,38090..38187,38291..38539,41319..41691))
/gene="FLN"
/codon_start=1
/evidence=experimental
/product="filamin"
/protein_id="AAA92644.1"
/db_xref="GI:1203969"
/translation="MSSSHSRAGQSAAGAPGGVDTDAEMPATEKDLAEDAPWKKI
QNTTRWCNEHLKCVSKRIANLQDLSGLRLIALLEVLISQKMKRKNQRTFQRM
QENVSVALFEIDRESIKLVSIDSKAIVDGNLKLGLLILHLISINPMDDEED
EBAKQTPKRLIGWIQNKLPQPIINFSDRQSGALGALVDSCAPGLCPDWSDWA
SKPVTNARAOQADWLGIPOVITPEEIVDPNDEHSVMTYLSQFPKAKLKPAPLR
KLNPKARAYGEGIPTCNMKKRAEFTVETRAGQGBVLVVEDPAGHQBAAKTA
NNDKFTFSVYKVPVETGTHKTVLPAGOHIAKSPVTVDSQSDASKVTAQGCLE
PGSNLANKTYFEIFRAGAGTGEVVIQDPWGQKTVBPQLEAGDSYRCSYQPTM
EGVHTVTFAGVPVPIPRSPYTVTVGQACNPSACRAGVGLQPGKVRKETAQFKVYTK
GAGSGELKTVTKGPEERVKQGLDGVYGEYPMVPGTVITVITWGQNGRISPF
EYKVGTECGNKVRAMQKLEGVGKGSADFVEAIGDDVDVTGLFVSEGVSAQIKED
DKDGSVDYRNPQEAQVAVHLCNSEDIRLSPFNADIPDOPHPDRVKARGPGL
EKTGVAVNKPAPETVDKAGGKAPLQVQDNEGCPVEALVDKNGNYTSCVPEKP
VHTGVNMGCVSIPNSPRNVNAGSHENKVKYVPGVAKGLKAHEFTYFDCAE
AQGDVSIIGIKCAPGVVGAEDIDFIIIRNDNFTVKTIPRGAAGSYTIMVLFAQQA
TPTSPTRVKEFSDASKVKAEGPLRTSGVELGKETHFTVNAKAGKGLDVQFSGL
TKGADVDDIIDHNDQNTVTKYTPVQGVGVNVTGGDPIPKSPFVAVSPSLDLS
KTKVSLGSKGVQVQKQDFTVKSAGGQGVKASKIVPGSAAVCPKVPGLGADNSV
VRLPREEGPYEVYTDGVVPGSPPLEAVAPTKSVKAFGPGQLQGSAGSPARF
TIDTKAGTGGGLGVVEGCEAQLECLNDGDTGCSVSYPTPEPDNINILFADTHIP
GSPFKAHVFCFSDASKVSGPLERATAGVQFQVDCSSAGSALNTEICEAAGLP
AEVYIQDHQDGHITTYIPLCPGATVTIKYGGQVFNPFPSKLOVFPVAVTSGVQCYG
PGLGEGVFREATFTESVLDCAQNTTGTGPHVKARVANPGLNLTETVYQDRGDGMKYVE
YTPYREGLHSVDVYDGSVPSPFPVPTVEGDCPSRVVRHVGPIQSGITNKPKFTV
ETRGATGGGLGLAVEGSPSAKMSCMNDKSGSCSVETPYEAGTYSLVNVTYGGHQVPS
```

```
Query Match 4.2%; Score 21; DB 9; Length 219447;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCCTCTGTCGTCGCTTCT 82
Db 99830 GCCTCTGTCGTCGCTTCT 99810

RESULT 23
BX640433
LOCUS DEFINITION BX640433 BX470249
VERSION BX640433.1 GI:33574176
KEYWORDS complete genome.
SOURCE Bordetella parapertussis
ORGANISM Bordetella parapertussis
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE 1
AUTHORS Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdeno-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N.,
```

```
Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Doggett,J., Feltwell,T., Goble,A., Hamlin,N., Hauser,H.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Norberczak,H.,
O'Neil,S., Ormond,D., Price,C., Rabinowitsch,E., Rutter,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skelton,J., Squares,K., Squares,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrell,B.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
Online Publication
Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227
REMARK 2 (bases 1 to 349305)
AUTHORS Sebahia,M.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: mssesanger.ac.uk
FEATURES
source
1..349305
/organism="Bordetella parapertussis"
/mol_type="genomic DNA"
/strain="12822"
/db_xref="taxon:519"
gene 110..562
/locus_tag="BPP3223"
CDS 110..562
/locus_tag="BPP3223"
/notes="Similar to Ralstonia solanacearum probable
transmembrane protein rsc0795 or r805042 SMALL:Q8Y195
(SMBL:ALU64061) (150 aa) fasta scores: E(): 6.4e-34,
63.3% id in 150 aa"
/misc_feature 1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="CAE38509.1"
/db_xref="GI:33574177"
/translation="MLQDALLAWLHYLAIFVILVMTABAVILRPMQMSARSVGRLLAY
DLRLASALAVLATGVLRLTGAKGAFTYMANPWFHAKTLFLVIALCSIPPTLAFLS
WKQSLSQPGFTPADDIRRARRWMIESHLFIPLFAVLMARGIGI"
join(122..190,248..307,350..418,485..553)
/locus_tag="BPP3223"
/notes="4 probable transmembrane helices predicted for
BPP3223 by TMHMM2.0 at aa 5-27, 47-66, 81-103 and 126-148"
gene complement(578..1729)
/locus_tag="BPP3224"
complement(578..1729)
CDS /locus_tag="BPP3224"
/notes="Ortholog of Bordetella pertussis (BX470248) BPP356"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="CAE38509.1"
/db_xref="GI:33574178"
/translation="MPAHLYGRILYHLRMPVLPVPSLSFVILLASGALALLGNSA
ALAFSGRRARAFRRRLACGLYAVLLVLAGYFAFGVAAVYRIEERWRHRTTL
PEARVGGIDMPAGTSLVLEAVSPDTYNRAEFHPVSAHGLQALRIERYIALDHPD
TYKLTEAYPLTWRTGTGTQAIQGWCDAAASPMFSTRDRIRAAFECEWLAENRAD
GMEVPPCARLLAHGTVYTDGHVDPDRWQVLPEDMAVRIGGVWLAIRLDAERY
DAFERAELACPLAFGPMHYPACTEVSAGRWERHEPGAWIFSLAGASARYAGHFDV
ADQAVVQGRGGEVLAVPNNRAGVLRFAIAVGNDAAPRAACPPR"
complement(1589..1729)
/locus_tag="BPP3224"
/notes="Signal peptide predicted for BPP3224 by SignalP 2.0
HMM (Signal peptide probability 0.997) with cleavage site
probability 0.556 between residues 47 and 48;
signal-peptide site"
misc_feature complement(join(1475..1540,1604..1669))
/locus_tag="BPP3224"
/notes="2 probable transmembrane helices predicted for
BPP3224 by TMHMM2.0 at aa 20-42 and 63-85"
gene complement(1779..3089)
/genes="gitA"
/locus_tag="BPP3225"
```

```

CDS
/note="synonyms: glut, icdB"
/complement(1779..3089)
/gene="glta"
/locus_tag="BPP3225"
/EC_number="2.3.3.1"
/note="ortholog of Bordetella pertussis (BX470248) BP2358"
/codon_start=1
/transl_table=11
/product="citrate synthase"
/protein_id="CAE38510.1"
/db_xref="GI:33574179"
/translacion="WNLSKATLSFSDGSAPIEPPVKGTVGPVIDIRKLYGTGM
FTDFGFSTACAGSIYIDGXGELLYRGYPIQLAVNCFDLICILINLGPNG
GQKQDFSOVTHVMNQLHFFLRGFRDHPMAVLGLVGLSAFVHSDTIDTNP
HRHVAIRLIAKHMFTLVAMAYKYSQQPIFYPNDLSTYGNFLRMVATPCREYKVD
VERALRDIILHADEONASTSTVRLCGSSGTNPAALAAAGVACLWGPAGANEAC
LQMLELANGVAKVGFMEKVKDNGSVRLMGFGRVYKNDYDRAKLMQBTCKEVL
VALGENDPLFLKAMELRIELEDYFVORKLPNVDFYSGIVQRAIGIPTSLSLTAIF
ALARTVGTQWNEMLSPDYKIGRPRQLFTGSVTRDVPQMDKR"
/complement(1836..2942)
/gene="glta"
/locus_tag="BPP3225"
/note="HMPfam hit to PF00285, Citrate synthase, score
2.2e-209"
/misc_feature
/complement(2124..2162)
/gene="glta"
/locus_tag="BPP3225"
/note="PS00480 Citrate synthase signature."
/complement(3110..3370)
/locus_tag="BPP3226"
/complement(3110..3370)
/locus_tag="BPP3226"
/note="ortholog of Bordetella pertussis (BX470248) BP2359"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE38511.1"
/db_xref="GI:33574180"
/translacion="MSRLTELQRLRWARRGLLENDLIIRYLDACEDQLTDBDVT
ALTQLFELGNDLLDLARKEGALDTPRLAIRDMREL"
/complement(3396..4112)
/gene="sdhB"
/locus_tag="BPP3227"
/complement(3396..4112)
/gene="sdhB"
/locus_tag="BPP3227"
/EC_number="1.3.99.1"
/note="ortholog of Bordetella pertussis (BX470248) BP2360"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase iron-sulfur protein"
/protein_id="CAE38512.1"
/db_xref="GI:33574181"
/translacion="WSTKRIVKFEIYRYDDPKDERPYMQKLEVELOPTDMLLDALVR
IKNDVDSLALRRSCREGVCGSDAMNGKGLACTNLLKLPVILKPLFGLFVIR
DLIVMTFFNYGHSIRYLIINDTPPEKERLQSPARELDGLCEILCACCSSTSCP
SFWNPNDFVGPAGLQXRYLFIADSRDEATGARLDNLEDPYRLFCHTINMVCVCPK
GLNPTKAIKIKELMVRVT"
/complement(3633..3668)
/gene="sdhB"
/locus_tag="BPP3227"
/note="PS00198 4Fe-4S ferredoxins, iron-sulfur binding
region signature."
/complement(3858..4049)
/gene="sdhB"
/locus_tag="BPP3227"
/note="HMPfam hit to PF00111, 2Fe-2S iron-sulfur cluster
binding domain, score 0.1"
/complement(4125..5903)
/gene="sdhA"
/locus_tag="BPP3228"
/complement(4125..5903)
/gene="sdhA"

```

```

/locus_tag="BPP3228"
/EC_number="1.3.99.1"
/note="ortholog of Bordetella pertussis (BX470248) BP2361"
/codon_start=1
/transl_table=11
/product="succinate dehydrogenase flavoprotein subunit"
/protein_id="CAE38513.1"
/db_xref="GI:33574182"
/translacion="WVAVMNSLPRQFDVVVVGAGAGMRCSLQLAQAGLSVAVLSKY
EPTRSHTVAAQGVUSASIGMSEDNWYHMYDTVAGSDWLGDDALTEPMCREAPNAVY
BLEHQFMPDRNADGTIYQRPFGHTANFGEKPVORACAADRTGHALHHTLYQRNVA
ARTQPFVWMLDLRLNAGDVLGTALMETGEIYVLEAKTIVLATGGAGRIWAAT
NAFTINGLGLMARAGIPLODMFEWQFHTGAGAGVLIITGVREGGILLKQGER
FMERTYAPTLKOLAPRDFVSRMDQFIKEGRGCGPDGSYVVLKDLHGLADVINKRLPSI
REIAFKGNVDPIKEPIVPVTHYOMGIPANYHGOVLERANGENKI VNGLYAIGEC
RAVSVHGARNLGTNLSLLDI VFGRTGNHIVDSHPEROHAQPLKEAVDRRARVDE
LESRTSGKTDQVANARMSMORHCGVFRTELLNEGVOI EDLQAQAHITFKDKSK
VNTARVLELANNTTEVARATIKSAANRTSRGAHLNDHPERDDVNWLKHHTLWYSE
GSRLDYKRVQMKLTVESFPFKARTF"
/complement(4131..4514)
/gene="sdhA"
/locus_tag="BPP3228"
/note="this hit extended beyond the end of the feature by
1 aa and was clipped.
HMPfam hit to PF02910, Fumarate reductase/succinate
dehydrogenase flavoprotein C-terminal domain, score
6.7e-61"
/misc_feature
/complement(4572..5567)
/gene="sdhA"
/locus_tag="BPP3228"
/note="HMPfam hit to PF00890, FAD binding domain, score
4.5e-191"
/misc_feature
/complement(5733..5762)
/gene="sdhA"
/locus_tag="BPP3228"
/note="PS00504 Fumarate reductase / succinate
dehydrogenase FAD-binding site."
/complement(5907..6290)
/gene="sdhD"
Query Match 4.2%; Score 21; DB 1; Length 349305;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 273 CGGCGTGGCGACCACTACCA 293
Db 138030 CGGCGTGGCGACCACTACCA 138050
RESULT 24
LOCUS S67307 5249 bp mRNA linear MAM 05-FEB-1994
DEFINITION Ca(2+)-sensing receptor [cattle, parathyroid gland, mRNA, 5249 nt].
ACCESSION S67307
VERSION S67307.1 GI:453108
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 5249)
Brown, E.M., Gamba, G., Riccardi, D., Lombardi, M., Butters, R.,
Kifor, O., Sun, A., Hediger, M.A., Lytton, J. and Hebert, S.C.,
Cloning and characterization of an extracellular Ca(2+)-sensing
receptor from bovine parathyroid
Nature 366 (6455), 575-580 (1993)
94077182
PUBMED 8255296
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 140593] from the original journal article.
This sequence comes from Fig. 2.
Location/Qualifiers
FEATURES

```

```

source
1. 5249
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
1. 5249
/gene="Ca(2+)-sensing receptor"
515. 3772
/gene="Ca(2+)-sensing receptor"
/notes="this sequence comes from Fig. 2"
/codon_start=1
/product="Ca(2+)-sensing receptor"
/protein_id="AA29171.1"
/db_xref="GI:453109"
/translation="WALYSCWILLAFSTWCTSAVGPORAKKGGDILILGLPPIHFG
VAVKQDLKSRPESVECIYRNGFRWLQAMIFATEEINSSPALLNMTLGYRIPTDC
NTVSKALEATLFSVQAKNIDSLNDEFNCSEHIFTIAVVGATSGISTAVANLGL
FYIPQVYASSRLSNKQFKSLRTIPNDEHQATAMADIIIEYFRNWNVTIAADDD
YGRPIKEFREAEERDICIQSELSISQYSDDEKIQQVEVITQNGKAGQVPGREF
LEPLIKEIVRRNITGRILASAWASSSLIAMPYFHVVGTTIGFLKAGQVPGREF
LOKVPKSHVNGFEKPEWETFNCHLQKAGKPLVDTFLGHGEGGARGLSNPTAF
RELCTGHEMISVETPYMYTHLRISYNYLVAVSIAHALQDIYTCIPGRGLETFNGSC
ADIKKVEAWVLKHLRHLNFTSNMGQVTFDECGDLAGNYSIINHLSPEDSGSVPEKE
VGYNYVYAKGDETLFNDKILWGSFREVFNCSRDCLAGTRKGIIEGEPCTCEP
VECPDEYSDETASACDQDDPDSNENHTSCIAKEIEFLSWTEPFGALTLFVLIG
IFLTAFLVGVFKFRNTPIVKA TNRLSYLLLFSLCCFSSSLFFIGEPQDWTCLRQ
PAGISFVLCISILVKTNRVLLVFEAKIPTSPHRKWGLNQLQFLVFLCTFMQIVIC
ALWNTAPSSYRNHELEBIEIFITCHEGSLMALFLIGYTCLLAALCFEFAKSRKL
PNFNKAFTITSMILFFIWIISFIPAYASTYGVKFSVAVEVIALAASGLLACIFPN
KYIILFKPSRNTAEVNSTAHAFKVAARATLRNSVRSQRSSLGSGTSGTSPSS
ISGKSNSEDPFQOQKQPOQLALSPHNAQPPQRPSTPQPOQSQPPRCQK
VIFGSGTVFSLSPDPOKTAVAHNSHTQTSLEAKNDALTKHALLPLOCGETDS
ELTSQBTGQGFVGHQLEMDPEMSPALVVSNSRSFVSGGSGTVTENMLRS"

ORIGIN
Query Match 4.0%; Score 20; DB 4; Length 5249;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 25
AR012622 AR012622 5275 bp DNA linear PAT 04-DEC-1998
LOCUS
DEFINITION Sequence 1 from patent US 5763569.
ACCESSION AR012622
VERSION AR012622.1 GI:3970612
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5763569-A 1 09-JUN-1998;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 26
AR012622 AR012622 5275 bp DNA linear PAT 04-DEC-1998
LOCUS
DEFINITION Sequence 1 from patent US 5763569.
ACCESSION AR012622
VERSION AR012622.1 GI:3970612
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5763569-A 1 09-JUN-1998;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 27
AR078217 AR078217 5275 bp DNA linear PAT 31-AUG-2000
LOCUS
DEFINITION Sequence 1 from patent US 5962314.
ACCESSION AR078217
VERSION AR078217.1 GI:10004963
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5962314-A 1 05-OCT-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 28
AR177781 AR177781 5275 bp DNA linear PAT 17-DEC-2001
LOCUS
DEFINITION Sequence 1 from patent US 6313146.
ACCESSION AR177781
VERSION AR177781.1 GI:17920136
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Van Wagenen,B.C., Balandrin,M.F., DelMar,E.G. and Nemeth,E.F.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 6313146-A 1 06-NOV-2001;
FEATURES
Location/Qualifiers

```

```

AR028465 AR028465 5275 bp DNA linear PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 1 from patent US 5858684.
ACCESSION AR028465
VERSION AR028465.1 GI:5940438
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Nemeth,E.F., Brown,E.M., Hebert,S.C., Garrett,J.E. Jr., Van
Wagenen,B.C., Balandrin,M.F. and Del Mar,E.G.
TITLE Method of screening calcium receptor-active molecules
JOURNAL Patent: US 5858684-A 1 12-JAN-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 27
AR078217 AR078217 5275 bp DNA linear PAT 31-AUG-2000
LOCUS
DEFINITION Sequence 1 from patent US 5962314.
ACCESSION AR078217
VERSION AR078217.1 GI:10004963
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Brown,E.M., Hebert,S.C. and Garrett,J.E. Jr.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 5962314-A 1 05-OCT-1999;
FEATURES
Location/Qualifiers
source
1. 5275
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 4.0%; Score 20; DB 6; Length 5275;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 GCGCGGTCTCGGCGCGG 381
Db 223 GCGCGGTCTCGGCGCGG 242

RESULT 28
AR177781 AR177781 5275 bp DNA linear PAT 17-DEC-2001
LOCUS
DEFINITION Sequence 1 from patent US 6313146.
ACCESSION AR177781
VERSION AR177781.1 GI:17920136
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 5275)
AUTHORS Van Wagenen,B.C., Balandrin,M.F., DelMar,E.G. and Nemeth,E.F.
TITLE Calcium receptor-active molecules
JOURNAL Patent: US 6313146-A 1 06-NOV-2001;
FEATURES
Location/Qualifiers

```





```
/product="luxR-family transcriptional regulator protein"
/protein_id="BAB69344.1"
/db_xref="GI:1582418"
/translation="MTVEDEQDEHDLGRAAPGDFELPDVSGFDLLQYRESLQGRR
PRETSQALLRHLLHVPSEASDVVCVPVSPSAADVALVAPLEERLKLKKINDIRSA
MAAFDAVHSEVKEAIVLADLVGGAAISASVTSASASEEIIITVQGGSRPPELLA
ETMELNLSLLAGRQRTLYQHSIRKKHQLTWDYGRQVVAAGAEIRTVTEIFERLICD
DTVAYIPNDRANEALEIRHPALIRYLARAGFERAWEAEPISPADPLPKAREDAIL
VRIQLLDVGHDTDESISQRLGIGVTVAKRISTLSGRLSRSRGQVLYIARSQWLDE
PSPAPPAPPSG"
CDS
1723..2700
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB69345.1"
/db_xref="GI:15824184"
/translation="MSSMWRPNPAELDKAVTGAERDAILDARSINDDARLGDSEAFY
AQQLLARIYGLHTLIPGPASAPSVLHITIRLLEAEATIAEDGTYIEPGLIDQPAE
PAAYLSWLKLAERHRYVYKHYREFIRDEATSDRLTYVQESVVDGRFDDLLAMQO
VGTSGAAKWEIAQNFDMGKGPKEVHTLNFKI FEVFEIPADELERSLTAANLLSG
NLAVLLCRYOYVPAVGLGTEWLVPDRFVQVVAWERLGLPDIGIVYHRLHITVD
SQAAGWPHNVVIPAASEHMRRAITARGTLMLNLSARYLDERMPAVAV"
2732..3427
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB69346.1"
/db_xref="GI:15824185"
/translation="MPLNEQFALLIEPVSQELSDSYEDLVDVYMGNGTRPKRFSQ
YRLREGTEGTFEALLAHRDYFTTKFNKVGQGIIRVQPIKADFPVLIISHGIEBGL
GRASEQINVHQNRADPGDVAGPLTPEGVHDHGHEFYMIAVLNKNVAGAQIRLWEP
DASDAPFWTGLTPKPGHAVLLDDRLGLAHDVTDESADGPGANRDIVIIAFSRWAERWYGE
BHDEAALADGAS"
3474..4397
/codon_start=1
/transl_table=11
/product="putative diaminopimelate epimerase"
/protein_id="BAB69347.1"
/db_xref="GI:15824186"
/translation="MGEHSTSTVTVTPASTPSAPPAPAEFRSABEFFTHALGNDYVV
VDPGSDFTPGPAAVRLLCORHFGVADGVLPGLPGAPAGPVRMLNSENDSGTCEK
SNGRLNFALHATYETGPEWSDERFPLTAAGDVPRVILDLASGTVQMGMRPFGA
TGESLDVGRRLAVTSVHNPNHVVPLAEISAGLARELGPLIAGHTRFPGRNTNVQFL
ALVDRRTAIEVYERGAETYLASGSACAAAHAALGLTGPDEVVRMPGGRIDVSLA
ADGVSMTCTAERISSGRSPVFRNKLETAR"
4432..5535
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="BAB69348.1"
/db_xref="GI:15824187"
/translation="MTTATVTSTVPMRRVVDADAGRYDIDLGDNSLPGCTILDELPL
GGLELYGNRGIPRLRALVADLYGTDDEHVLITHGAQQAQLVLYNALLRPGDOVITF
RPGWQSDMPALGCRVDILDRPOLTVDEAAAVAGPDLRLIYVNTFCPTGQVR
SDADLAALHTLAARTGAYLLDDEYALKELOSAATLERITYSVSSLSKVGYPGLRIG
WLYGPPLVACTTRELGNISNVLCTIACDVLERRDRHYTERYLTGEGLVAVRA
WAARNDVAVLTPEPTGPPAWFRLLTGEAPLDLCRRVLDGTGVLIMPGETILGGDGRFL
CFARDPHLISEGLRRLLDPVLRPOAVENMTSGAV"
5673..8723
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69349.1"
/db_xref="GI:15824188"
/translation="MAYRLRGTLDVGAHRLHSLVTRNTALRTTFRSAGTYPGVVA
PAGEVLEIVDLADADLTRARSEAPFDLAAGPLYRFLSLRSLAEDHVLVIVFHHI
NTDQWSALINRISISAYNAFVQGTQPALPELATYEDHVRARQEQGGDSWQQLKF
WKCKLAGLVVEMPADRAPATPTONGETILTFDPAELLHGARALAQEGASFLMVLIT
TALNVVGRVTGQEDIPGLIPMLGRTEPELEDVYVGLFINMLVRLTDLSDGPTFAELTE
RVADANMVDYNDQDFQIVSHVQAPRAGNPLFOIGTQLLGGSGTGGSDLGTLGLT
DEQDLVAPSRVLDLSFLESADRLRFVAEYITDLDGIRIEMARHIONVLAAYE
DAPQPI SKLAMLGEDERLLEAGRPDFFSHDPVHTTVLQVAERTPHAVAAVCRGV
EMTYGELARRSGKLARHLRARGIRPGQVVGIVMDRDLDAVLGVNRAGGAVAMDPA
```

```
SHPAARDYMLQDTAAVVLTRADVAGGLPTSDEDRKVVLLIDADWAAIEAESDTPELEE
LADRETIAYVLIYTSGSTGRPKGVMIERSLAVFIEAYRHTCEGETPDRLLQOSSLIF
MCSQETIYALITGATIVLSPDEASSPEALGILIRGQRYVYVGLSPTLLSLLGGPY
PHLYXIMGGAEEVLPABLNVKNWIPGRFVNLYGPTAAIACTEYLCEHIEWKSPFPG
RPEYGRHYVVDHRLNPAVPGGELLIGEDGLAGLYNLPELTDEKFPNPFPEPTG
RVYETGLVWTPDYQVDFIGRIQNVKRLGLRIELGIESALVAHPQIRMALVLLKA
DPQEKOLGVYTVLPOQAPTLAELRHLGEMPEYMPVPTAWVKLDBFPPLTAKKIDR
HALPEPSYGADCTEGRAYVAPATTESERVAEAFATVLKADQISADASFFELGGNSLQA
MAYRSRINKAPGVKINVMLYGGATVGVGAEDIDELLAAKAGRD"
8737..9549
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69350.1"
/db_xref="GI:15824189"
/translation="MPLREDGTHPLYCVHPVSGSYVYAGLAELLGGQQSYSGEAF
GFDDGAPLNSIEALAEHRLASLRPRGPYLLIGSLGGVAVYHMAQRIVAEAGEV
PLVILIDATVPQPPPLPEKALHAFRLDLVTAMGEPPEALDALITPYDDVTPAVV
TAVTAAGDEVLPSEDEEFLOERYDVFEAAEALVAYERHGFYAGRVLPFLKAADSLDA
RTRMETLAPRVERHITITGDHDSIWAGEAFATLAGLVROQLAQAARSALAPASG"
9609..10901
/codon_start=1
/transl_table=11
/product="putative export protein"
/protein_id="BAB69351.1"
/db_xref="GI:15824190"
/translation="MEDTAAAGSAAAPRRGGGLWRNRDPLVFWTGETLSLYGTQVTL
ALPLAAVFDASAEELGVRLFQLVFPFLGLAMFGVMVDRVKRPPVMIGANAARWL
IGLVPALGAADVALLVAVAGVQLVFPDVSWSYVPLVKDRHRLVANAATKLS
TSSSADVAGPGVAGVYVNLATPATAVNAVYLLSIVSLIAITPPEPPAKPRRR
STRADVREGLRPFVGSRLRALALVGCACNPFVSMLLSLFLVAVRGHLSAATIGLV
LGVAVGGIAGSALSGTVIRRVSVGKAYLVSVSGIFVAPLLIPAGGSKLWTVGFIA
GLHEALWLAASVAVMLVPLIALSPVSRLEHMPAPAPVP"
10979..12856
/codon_start=1
/transl_table=11
/product="non-ribosomal peptide synthetase"
/protein_id="BAB69352.1"
/db_xref="GI:15824191"
/translation="MSPDVEIAAAMWGSVPHYDEIAAMNDTARELPGGGDTLPELPR
QVSRPTDPAVFRDTRLSYTELDSRLARLLAGHGAAPDRVVALTVPRSVELVVA
LNAVTKTAAVLPDPDPAADRGVLYLADAAVLLITHSDVATLPPGAPGELPLRV
VDAQSEPLERAPFAPPPAADAAYVITSGTGPVGVVPHAGVNLHWMQHEY
RUTGDVLOKTPGSDVSVWBFPLITGATLVVAEPGKHDPALYLAELIRGERLIT
VHFVSMLOVFLDEFAALCTGLRVVCSGEALPVELADRFTASLPGELVAGVYLGTE
ASVDVYETPEPGATSVPIGRPYWNTLVYLDALRPAAVGETGELVAGVYDLALG
VYNRPGLTAERFADVPYPPGRMWRTGLARRRPGDGLDYVGRADHVKIRGLRVEL
GELEALABHPAVSAAVLRDRPGDKRLVGVVPAVPGTVPABRLTAAATLTPDY
MVFSTVVVLESPLTPNGKLDKALPAPDAAAAPARVAPRTPQEALLCDLFAEVGL
SETQVGVDSFLELGGHLSAURLVSRIRAAFTDCLKVRQVFTPTVVAALACLASNG
R"
complement(13245..14198)
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein BldK"
/protein_id="BAB69353.1"
/db_xref="GI:15824192"
/translation="MQAVDGLDFQVAEGESLGVSGSGCKSTTGRLITRLLPFTSGK
ISYRGDITHADRKLPVRSEIQMTPODPYASINPRHTVTKIAGPMEINGIQPSGG
REKVRRELLIEVGLNEPHYRFPHEPSGGORIGIYARALALEPKLIIVADPVSALDV
SIOQVNVLLIQDLQKOLGIAFVFIADHLALVRHFSORVAVMYLGLKIVEIADRADLYEN
PRPHYTRALLSAVFEATVEDGEAREIRLITGDVPSIPSGGCRFRTRCWMKATKCMAS
EAPLVQVGVGNAGHLTACHYPTKETVPAPRLSKDPPKAAV"
complement(14315..15385)
/codon_start=1
/transl_table=11
/product="ABC transporter intracellular ATPase BldK"
/protein_id="BAB69354.1"
/db_xref="GI:15824193"
/translation="MKSTATKPAERNPQATGPFVSQDLKVHFSTBDGTVKAVDGLS
FDLBERKTLIGVSGSGSKSVNTLITGLNPNDSATVEGSIITLDBELNGASEKTEK
LRGNKMAWIFQDALTSLSPHTTGKQIGETRYKHTGSKAAARALEMTLRVGIPOF
```

```
DVRVDYDYPHQSGMQRAMIALVCDPELLADPTTALDVTVOAQIMDLIKOLQK
EFTALIFTHDGLVADIADDDVIMYGGRCVERGCKEVLRTPOHQHYTWGLIGSMPS
LEGFVDVPLTPIPGSPSPLNPPSGCRFRCFAEKVEDCRSSRPPLLEITDGRGA
ACHLTADORRAFFADLAASPAN"
complement (15446..16339)
/codon_start=1
/transl_table=11
/product="transport membrane protein BldKC"
/protein_id="BAB69355.1"
/db_xref="GI:15824194"
/translations="MPRDPARLACGKLCPTPETLEVRHNNMTDSDVQVWHMLTGIF
LGRDYSYGHCDAPCLGSGVNRRESVLGTTMDRPTFTASLAGSVVFLGGICGAMI
AAVKGQKALDPCAGSGLVASSMCIYFVGTALYCFYQWQILDPQSYTDFTKGNASW
ANGILLPWLVLISLITWANYTMRRESOLVNEQYVTRAKGLSRNRFPRFARWGA
VPIVITVDYDGLGVLLGCMITEKTFSLQIGMLAKAVTNNDLPLMLLGIMVITASFV
MGIANVVDAYVAILDIPRILIA"
complement (16552..18315)
/codon_start=1
/transl_table=11

Query Match      4.0%; Score 20; DB 1; Length 23451;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      22  ACCGGCGGCCCAAGGCGGT 41
Db      11558  ACCGGCGGCCCAAGGCGGT 11577

RESULT 32
AF512431      28890 bp      DNA      linear      BCT 04-JUN-2002
LOCUS      Saccharothrix mutabilis subsp. capreolus nonribosomal peptide
DEFINITION      synthetase gene cluster, partial sequence.
ACCESSION      AF512431
VERSION      AF512431.1  GI:21321595
KEYWORDS      Saccharothrix mutabilis subsp. capreolus
SOURCE      Saccharothrix mutabilis subsp. capreolus
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardiaceae; Actinosynnemataceae; Saccharothrix.
1 (bases 1 to 28890)
Yin,X., O'Hare,T., Gould,S.J. and Zabriskie,T.M.
Molecular Cloning and Sequence of a NRPS Gene Cluster from
Streptomyces Capreolus Az50
Unpublished
2 (bases 1 to 28890)
Yin,X., O'Hare,T., Gould,S.J. and Zabriskie,T.M.
Direct Submission
Submitted (13-MAY-2002) Pharmaceutical Sciences, College of
Pharmacy, Pharmacy Building Room 203, Corvallis, OR 97331-3507, USA
Location/Qualifiers
1..28890
/organism="Saccharothrix mutabilis subsp. capreolus"
/mol_type="genomic DNA"
/strain="Az50"
/sub_species="capreolus"
/db_xref="taxon:66854"
<1..728
/notes="ORF1; similar to penicillin-binding protein 4"
/codon_start=3
/transl_table=11
/product="unknown protein"
/protein_id="AAM47271.1"
/db_xref="GI:21321596"
/translations="HAWNALKRVPVVRQSLAVDAPAGALASATDLVTFGRMLAGA
PPTLPDLDLDMRAVPVRAEPFGMDGGLALVERDGAVMVGHGDTAGTGSCHLR
INFDDGTVALVTNGSGGFRALWDLVPELAAGLVPADYDGFALRRRIIPPEPCAGR
FRNGDTYKVRPADCEBELITVDGEPADLSLFDGLVFMKDSVTGTSQTGRFLRDP
QRGSIEWIQIGRLARRQVPMV"
878..8371
/gene="scpsA"
878..8371
CDS
```

```
/gene="scpsA"
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="peptide synthetase ScpsA"
/protein_id="AAM47272.1"
/db_xref="GI:21321597"
/translations="MVCAEGAVSYRELDARANKLAHLILLAKGARPERVVAIALPRSDV
NVVARLAVLKGTGAAYLPVDFPAYPAERIDFMLEDAEPLLVSTAPGHRRAAHRRCRCDP
PRLPAAVITSGSTGRKGVVTHRGLEPAFAEIAHFVDRPGDVRVQFSFSDASV
LELMACAGAAALVVPVPGPLICEQLADVIDGVTGTHALIPVVALATVPDRPIPTERC
LVVAGACAPADLVARWAPGRTWYNADVTSTVTSWSPLEPGGAPPIGRIPIGTVE
RVUDDLEPAREGELIVTUGLARGYLRRRGLTAARFADVPFGPGSMVRTGDVVR
APDGEALFVGRADHQVKIRFVRPEPEVEALLREHPAQRDAVPVVADEPDELVAPRT
EVEGLREYLAATLPDYLPVPSAFVLDPTPLSPNGKLDKRDALPVPVADPTEDEFPAPRT
DAARRVAQAOEVLGVLGAGHDDFFALGDSILAVRALRGLGSLFVAMFEHRTVSA
LAALVLTPTPIPRVPHDRPLPSPAQRCLFSLDGTAEQNTGVLRTGPTLDVSRLEAA
LDLARHARALRTEDVVDGEPQVVDADHGTIPLVVGDERELCAPFLDRRGPLTRAVL
KPLGPDQHVLIIVQHHLITDQVGSVANVLADLADYAGVEPAAPAOYDEPAVMDADR
HQDAAYVVERLAGLEALDPTDRPRPALRTAGAVHRRPLPRDLVRLTAVGRAHDAT
LFTTLTAAVQLLSARARQDIAGVTVTSGRDHADLERAVGFFVRLIVLRSWDDPDL
FTAPLDQVGTGTVVEAFEDHVPFRLVABELGVEPSPRTPLVQAVVALQPLVERADP
GGALAEHDLPHPHARFDLVVEFWPRDSDSLTIVYNTDLFAETVEERLAEDRLRLH
DVVAHPDELSLVELDPDDDAVRVGLAVOLAEBETLKRHHEVADAVVVDGRL
VAVVAPPVPAALKFGAQILPAHAIPTATFVGLHRLDADLPAPPRSDGVVAPRTPT
EAVLADFAEVLGAPKVGVRDNPFFALGDSILGILQVTRARAAGLIVLTSRDLFAHQTV
AALAPHVTRDLPDADQGPVTGDAPLTPIQRMFLGTGSRULPEQLSLVDLDDFAVTEADL
ADVAALVEHDLALRMEFEADRVNRPVGGPDLRLGELGPRSVRLTAHLVVDVVS
WRIIAEDLVTLALRRPELGRKTTSYLDWARRLAGHTRAGGFDDELAHAAVTCGVPTD
RGANTYATREVTVRLSEEDTALLRDVPEVTVQNDVLLTALGRVLADWTGRDVR
VVDGHRGRELFPADILSRVGMFTSVVPVALDVPREMGALKSVKERLRAVPRGI
GYGLRIAGGACDPAVGFNYITGWGGQKDLDTADPMPRPHLLDVGVGDGRL
EVAMVAAVHDESTVARLAESAAYAAALRAVTAHCAAPGAGRTPSDFELAKLTQEQVD
RITHYGAAPLTPMQAGWVHGLGERGLGSDARVYFQIITFLVDGPDSSAFARAWQ
RVNDRTPVLRSSVLWEDVPELPQVHHEAEVFTFLDMSDLDEADARLERVLEQDR
ALGIDLSVPLRMVAVVLAPELVTVVTHHLLDQWSVQVSLDGLGEPAERPPFR
DIWAHLAQDQAAERHWRVLTGFTDSTPLPADPAPAGPAASRRHAAVLSABSDR
LYAFARHGLTNTVMQAAWALLARSQRDEHVEPLTRLKAWSAVADGVLDLSDSAVVP
TLPVRAVDGTRPVVAMRLQAAQASRRFERHVEPLTRLKAWSAVADGVLDLSDSAVVP
ENYPVELGDGRGIEAEVETTSFPLSATVPSQRLGVLGVPAPVAPAAVTERUGER
LKALLVGAEDPERRVAVPMLSDAERQQLVDWNGDAEFFVTTTIPERFAQAARTP
DAIVSFAGEHLTVRELDERSNRLAHVLAAGAPERLVALLFETLADLVANVLGLK
SGAAYLPDPAVPAERITAGTADKPVVRLDPLDSCYPTTAPRVELHPENSAVVI
TSGSLPKPGVVIHNSNVIRLFAATRWVHGFEDENDVWELFHSYADSVSWELWGLLH
GRLVVSYETSRSRPERFARVREGVTVPNQTSFAYQLIPEQDESERVIFAGBPLD
LDKTDWKGAGTMNMGITVETTVYTVADQSGVPMPLRVYVLDDEDEPEVPPV
VGMVAGPGARGVLDRLGPTASRFVAVVLAVERLVAVVYVPGHVAVADLRHGA
DHQKIRGFVRLGEIEVLAHPAAVAVVLAVERLVAVVYVPGHVAVADLRHGA
AALPDYVPAAYVALDRLPLNANGKLDRAALPAPERDAVTSTYVAPRTETRAAEI
WAEVLVDVRVGVDESFALGDSIRSLRITSRITKAAFVDVLSPRDVLARTVSHLAGL
VEELVLADLEALADREA"
8375..26956
/gene="scpsB"
8375..26956
/gene="scpsB"
/notes="ORF3"
/codon_start=1
/transl_table=11
/product="peptide synthetase ScpsB"
/protein_id="AAM47273.1"
/db_xref="GI:21321598"
/translations="MTTISREDRISALPAHLRELKRLAGRAAPADVIIPRAGHDGPPF
LSAQRRMLFHLQVAFPTDYNTALALRGELVDRLTAALRELVRHRESLRTTFPEVD
GSPAQVVRDRLDVPVVDADDLDAALRAEVRFPDLADGVFVALVIRHAGDHLVLLT
SHIIVDGSILGVLTEELGALYGEALPPTDLRYGYDVAVQQRPADLDYWRDLAGV
TLPDLTPDRPRAERSAGATETVPPLASRLAEARLTETTLTLLTACQALLA
RYTQDDIANGVITVTCORNPTELQRLVGFVGTIVLSTVDRTARSFADLLTAGMGVLD
AFADHAPDKVVEALVPRDPRPTDFDVMVMQAGTALPALPGLYTEHEHLTRRH
ANWLDSPFAETGDGIGVVEYSTDLDFAATVARLCRHLILILARVVAEPDELAEPV
LILGDSFSPGPAAEVPTTFPALFAQAARTPDATLVCGDDVLTYAEAEANRIA
HRLVADGVPEQVVALSLPGAVVAVLGVKAGAVYLPVDPSLAGAACQACGHLG
RGAGRPGGHDPRPASHAAGPQGTGVRVYVTSGSTGTPGVVHRNLVNLVNH
RNDFASQPLKVALTVFSFDTSEGLVMADGHEHLVLTDEVALTEPTALVAVVREHRI
```

gene

CDS







```
exon      4413..4481
/gene="B10K17.010"
/number=1
intron    4482..4557
/gene="B10K17.010"
/number=1
exon      4558..5073
/gene="B10K17.010"
/number=2
gene       5386..5959
/gene="B10K17.020"
complement (join(5386..5508,5566..5627,5680..5959))
/gene="B10K17.020"
complement (join(5386..5508,5566..5627,5680..5959))
/gene="B10K17.020"
/number=2
/notes="contains EST gb:AI320151, AW722516
match of 4 ESTs"
/codon_start=1
/product="putative protein"
/protein_id="CAE75710.1"
/db_xref="GI:38524245"
/translation="MERTAIRCAOKASTSTANSSELPOAFVKLPHISPALLTOYPA
KATWPPDKTMSPOQLKYEKYEKKIHHIAQRPNNKMWQLAOLGTISFVLIYCLLF
ADWKDERQPFYEFQWFNSLGFYEAPKPVARIQSRSQASVQGSRRQM"
complement (5386..5508)
/gene="B10K17.020"
/number=1
intron    complement (5509..5565)
/gene="B10K17.020"
/number=1
complement (5566..5627)
/gene="B10K17.020"
/number=2
complement (5628..5679)
/gene="B10K17.020"
/number=2
complement (5680..5959)
/gene="B10K17.020"
/number=3
gene       6783..10216
/gene="B10K17.030"
join(6783..9476,9539..10216)
/gene="B10K17.030"
/notes="contains EST gb:AI392195, AI392192"
/codon_start=1
/product="leucine--tRNA ligase, cytosolic"
/protein_id="CAE75711.1"
/db_xref="GI:38524246"
/translation="MADTAAVAKGVNLSVSAKTKELKTEKRDITLIEIKRYQKW
EQGVFVDPAPSAEPFLDAITPDELROKHPKFFGTIAYPMNRLHAGHAFSKIE
YHTGFARMQGRALFPQGYHCTGLPTIKASADKLKTEIMFGQEFERYKEDEVVEGAAP
AAAAKPTKEDLTKEAKKGKTVAKTGAKYQOILKSLGI PVSEIHKPADPOYWLHY
FPCEKDLTNFGARIDWRROQFVTTDANPYDAFVWQNNRLLELNKIKFGKRYTYS
IKDQPCMDHDSRGGVLPQETALKLKTWEPAPKAAEALKKLEGANVILCPATL
RPEMTYQVCCFPALQGVFKAAENEYFVITERAKNWAYQGIKEKEGVISKAADI
VSDLTICTLVNKHYSVHKYVYVLPMDTIVLATKGTGVTVSPSDPDCAMMTLAKRP
EFGYIKOWEAKIEIVSKTPTSDLLAPLVKLLKINSPKADQLLEKELAYKEGFY
QGLMYGDRGEKVEAKPKVROOLDAGDAFAYSEPNKWSRSGDCSVALMDOWY
IDGERSWRVLYDYENKDKGINTYADTOHAFKVGICWLKQWACARTYGLGSKLP
WDPNFIVESLDSITVYNAIYTVVAHLHLRDLFGREKKGNGICADQMDVWDYIFCRTE
LSHLVTKGIPKETLDSMRREFQYFYPDLIRVSGKDLIPNHLITFWDINHIALFPREY
WPKSVANGHLQNGKMSKTGNFTMLDDVVKYGDAAARVALADAGDGSIDSNFVE
DVADNTITLRYTNKWEETLTKDSELTGIBNSFDALFDNMNVALNEARKHYETS
YKLAKAAHYDLFNARDMYREACAAAGIPLHKDLVTKYIRLOALVITPIAPHWADYVW
QECLEPKGSQFNARWPEVPAANPALTIARDYVITTSAINSAEAOCLKWAKGROSDP
DKPKPKLTIATENETWQAKYIDLLSEVWDAAATGTQKIDDKELNGRIAMGEMKKA
MPFYQALKRLKDGEPAEQILSRKLSFDEKATLIAMTIPGLKRTAGLEVSQVVLVEGGS
KTGKDLTNGGAETEVAPMAEALPGQPSFFFTNV"
/number=1
exon      6783..9476
/gene="B10K17.030"
intron    9477..9538
/gene="B10K17.030"
```

```
exon      9539..10216
/gene="B10K17.030"
/number=2
gene       11338..11931
/gene="B10K17.040"
join(11338..11443,11519..11931)
/gene="B10K17.040"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAE75712.1"
/db_xref="GI:38524247"
/translation="MRSSTLLAVLGTASLAQAOTITSYVSTPTSTSLTPPRONERL
QSCLAACMIQFPALPRVTGGGVIPPPAPAKGCRGRTRDLDRVADVGPCHPQPSPP
LLPVTFFPPFPKAPIQIOTPRPVFVNSLSPDPSAGFVPGNEDENQOTYEDPFLPD
DPPVASLPLQGF"
/number=1
exon      11338..11443
/gene="B10K17.040"
/number=1
intron    11444..11518
/gene="B10K17.040"
/number=1
exon      11519..11931
/gene="B10K17.040"
/number=2
tRNA       13251..13322
/product="tRNA-Trp"
/notes="tRNA predict as a tRNA- Trp : anticodon oca"
13933..14292
/gene="B10K17.050"
join(13933..14127,14239..14292)
/gene="B10K17.050"
/notes="similarity to hypothetical protein SPAC8F11.02c,
Schizosaccharomyces pombe, PIR:T39177"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAE75713.1"
/db_xref="GI:38524248"
/translation="MSDEYSIYDEVEIDMTYDPALQYISYPCPCGDKFEALADLQ
DQDIAVCPSCSLMVRVIFVDNLPKAPDAAASIAVSA"
13933..14127
/gene="B10K17.050"
/number=1
intron    14128..14238
/gene="B10K17.050"
/number=1
exon      14239..14292
/gene="B10K17.050"
/number=2
gene       15505..17227
/gene="B10K17.060"
complement (join(15505..16985,17047..17227))
/gene="B10K17.060"
complement (join(15505..16985,17047..17227))
/notes="similarity to hypothetical protein YDR132c,
Saccharomyces cerevisiae, PIR:S51859
contains EST gb:BG278421"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAE75714.1"
/db_xref="GI:38524249"
/translation="MTSSGHGVPARNGEHGVSVISASSHORIVHINILPHERVFPQ
IGTLEFKLGASLSDAPSYFSQYFLCQKQADAGQDPTAIRTYIDKDPVTFKDI
SLHLOQYHVRPRDATHLRLFADAQYSLPKMSQLYEETIHTIGSTPQIPRELI
LSGPGTNPFTTIGAFMFFNSPDLPGLDKEGLIRPPSIVAPNPNRSAEVQELIH
MLRGYPVKIDDERHQELKDCWVNFKLEQRLIPEHKGNOVRGREETVLGRDIM
KSGISVANEPVEGDPLAAMVNYNRYADDKAMELVLEIGREATQLHIGPGKEVRAE
LDPEASITLDGPQTADESDTAHPSLPANGTNTTPTGTTIPASASSSGQPSAETPD
FTTDIIRAKWAKLFEVIAATKLNLPFTQPSGLMNAKGGSEPTSRDALLUSEDIVRVI
SINPSTKPSFPQSSSPDNPNPKRRRLDSSPPYGVTSVTSPLSGEQWIVKTQW
RLFIOSSPNENPAKASVVECCLVAVKLDALSSERKNERRGFLGG"
complement (15505..16985)
```

exon

[illegible]



```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD16317.1"
/db_xref="GI:17429632"
/db_xref="GOA:Q8XW65"
/db_xref="SPTREMBL:Q8XW65"
/translation="MAKYSIAKALKTKVRKAGLSQEAFFSDVSSRTYMSLTLELDLSP
TLHKUTLCEVMDVHPLFILLTLAFAGDSTRKADELLAQVRQLEAVLNSDGD"
3522..4085
/gene="RSC2611"
/locus_tag="RSC2611"
/feature="synonym: RS00860"
3522..4085
/gene="RSC2611"
/function="miscellaneous; unknown"
/feature="synonym: RS00912"
Gene name confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAD16318.1"
/db_xref="GI:17429633"
/db_xref="SPTREMBL:Q8XW64"
/translation="MAYRNGYTAFFVAREPHPSLGNATKDFQYNTLRLMWKGADA
TPFVDSHKTYSVRGSDWEITLKLRLERLSSKNILILSSITNSRALREEDYD
GINDQGLPVIVYPTATKESLLVDSGLQAVKNLGNLVPFRDSMSKPTLTHVPNKK
ELIKSLDQGEFVMSGKAPDIYRNP"
4107..4955
/gene="RSC2612"
/locus_tag="RSC2612"
/feature="synonym: RS00912"
4107..4955
/gene="RSC2612"
/function="miscellaneous; unknown"
/feature="synonym: RS00913"
Gene name confidence : hypothetical
Gene name confidence : hypothetical
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD16319.1"
/db_xref="GI:17429634"
/db_xref="SPTREMBL:Q8XW63"
/translation="MIHMAKVNFDRVIRKKELEITAVVSGTSLSFIVFVDPAEWK
LKAGLAFALITLVYVIMWSNLSIDINVGSDVAIKTGDIFQOGLKIAFNBY
FDQVDNKKIIGETSLNGVFIQKHLGVPISLDRIEYAFDESEVLEENARKYKKK
RYOIGTICLYKDYLTAFSKFDENKALLTWPYLEFLINFDKVNNVVAQSVSTTI
FGSGITRIKGHKNSDEDLKIMLTFRISMRKYPKALTIVHKDKIDQINLLDIK
SARNGV"
complement (5305..5619)
/gene="RSC2613"
/locus_tag="RSC2613"
/feature="synonym: RS00913"
complement (5305..5619)

gene
CDS
Query Match 4.0%; Score 20; DB 1; Length 203050;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 CCTTCCTGGCGGCGTGCCG 282
|||||
Db 20919 CCTTCCTGGCGGCGTGCCG 20900

RESULT 36
AE016925/c
LOCUS 208524 bp DNA linear BCT 14-OCT-2003
DEFINITION Chromobacterium violaceum ATCC 12472 section 16 of 16 of the
complete genome.
ACCESSION AE016925 AE016825

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AE016925.1 GI:34105514

Chromobacterium violaceum ATCC 12472  
Chromobacterium violaceum ATCC 12472  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Chromobacterium.

1 (bases 1 to 208524)

REFERENCE  
AUTHORS  
CONSTRM  
TITLE

Brazilian National Genome Project Consortium  
The complete genome sequence of Chromobacterium violaceum reveals  
remarkable and exploitable bacterial adaptability

Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)

22882880

14500782

REFERENCE  
AUTHORS

2 (bases 1 to 208524)  
Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de  
Almeida, L.G.P., de Almeida, R., Goncalves, J.A.A., Andrade, E.M.,  
Antonio, R.V., Araripe, J., de Araujo, M.F.F., Filho, S.A., Azevedo, V.,  
Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C.,  
Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A.,  
Brocchi, M., Burity, H.A., Camargo, A., Cardoso, D.D.P.,  
Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M.,  
Falcão, C.L., Fantinatti, P., Farias, I.P., Felipe, M.S.S.,  
Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,  
Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A.,  
Goncalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grisard, E.C.,  
Guimaraes, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J.,  
Leoi, L.C.T., Passarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P.,  
Macedo, M., Madeira, H.M.F., Manfio, G.P., Maranhao, A.O.,  
Martins, W.S., di Mauro, S.M.Z., de Medeiros, S.R.B., Meisner, R.V.,  
Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F.,  
Oliveira, J.G., Oliveira, S.C., Paixao, R.F.C., Parente, J.A.,  
Pedrosa, F.O., Pena, S.J.D., Perreira, J.O., Perreira, M.,  
Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Potrich, D.P.,  
Neto, C.E.R., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos  
Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Senanez, H.N.,  
Silva, A.M.R., da Silva, A.D.C., Silva, D.W., Silva, R., Simoes, I.C.,  
Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, K.R.L.,  
Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R.,  
Urmey, T., Vettore, A., Wassem, R., Zaha, A. and Simpson, A.J.G.

TITLE  
JOURNAL

Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de  
Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ  
25651070, Brazil

FEATURES  
source

1. 208524  
Location/Qualifiers  
/organism="Chromobacterium violaceum ATCC 12472"  
/mol\_type="genomic DNA"  
/strain="ATCC 12472"  
/db\_xref="taxon:243365"  
complement (96..563)  
/gene="pile2"  
/locus\_tag="CV4209"  
complement (96..563)  
/gene="pile2"  
/locus\_tag="CV4209"  
/note="identified by sequence similarity; putative; ORF  
located using GeneMark/Blastx/COG2165"  
/codon\_start=1  
/transl\_table=11  
/product="type IV pilin"  
/protein\_id="AAQ61869.1"  
/db\_xref="GI:34105515"  
/translation="MKKQKQOQFTLIELMIWVIGLAAIAPAYODYTKRVRVSE  
GLADAARAAKTAATVEYATNNASGTPENTTYGLAASITGNSSVSNLAVGVQIA  
INSTVNSGALVDLIVPTVSAGSVQWYCTTSGTTLAASASQLSANWPTCRQ"  
838..2184  
/gene="bioA"  
/locus\_tag="CV4210"  
838..2184  
/gene="bioA"  
/locus\_tag="CV4210"

gene

CDS

gene

CDS





```
/codon_start=1
/transl_table=11
/product="putative two-component system response
regulator"
/protein_id="BAC72393.1"
/db_xref="GI:29608339"
/translation="MAIRVLMVDVDDQVLLRTGFRMLAAQPDMEVVAAGDGVAEALQVL
RSTVDVLMVMPKLDGVETTRICSEPNPKVLIITFDLDEYAFSGLKAGASGF
MLKDPVPEGLLTATRSVHSGDAVAPSTRELLDRFAPMLPSACKGPPKLEBLRTDR
EREVNLVAQGLSNGEIAARLVLSAEATVKTHVGRLLIKLGLRDRVQVVVLAYETGLVR
AGGQT"
8125..8757
/note="SAV4682"
/codon_start=1
/transl_table=11
/product="putative TmrB-like protein"
/protein_id="BAC72394.1"
/db_xref="GI:29608340"
/translation="MLLWINGPFGGKTKTQTAHEIQRRLPGSVICDPEHVGFLRRMLP
PGLRANFOLASWQGVVEVLDAIRHDGVVIAPMTITDSCFAQTGRLRELGHDV
RHEALLARETVLRLRERGRGHLLRYVAGKADAPLRGSAWAGQLDHCLERLRBPEFG
EHLWTDQSTVTKADRLAVAGLITLAPNTEGALRRLRQVRVGVKMRFD"
complement(8764..9888)
/note="SAV4683"
/codon_start=1
/transl_table=11
/product="putative AtrE-family transcriptional regulator"
/protein_id="BAC72395.1"
/db_xref="GI:29608341"
/translation="MSVCIDITGLPAERIVETSPLEIAVALHALAPGHHPATQGW
ATATSAALPLDLASQALFWSFTSDIFMFAGVAGVGRPGATIAEDLIDRLR
DDEEFTVAALPESQSGVAAGLPSPLVARHAIALAAASRQGLQRTORLEADPP
TVARVRLLEDCEAFADTWRVRVHLPOLADARHKTELLRHKGLREALTSVAALS
NTAGRTLEKLAHGAATADPAVGGVETVPSVHGMPHLLVHAGSWRPVHVVP
AAPEPSAASVELLKLRLALHPVRLRCLRYIARAPYTTGELAHADLSAPEVSRHL
TVLKAGLLTTRRGYVHLQDVTVAKLGSDLEGLIR"
complement(10130..10318)
/note="SAV4684"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC72396.1"
/db_xref="GI:29608342"
Query Match 4.0%; Score 20; DB 1; Length 299800;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GCGGCTGATCGAGCGGAG 256
|||||
Db 136409 GCGGCTGATCGAGCGGAG 136428

RESULT 38
AP005939
LOCUS
DEFINITION AP005939 300900 bp DNA linear BCT 12-SEP-2003
5/31
ACCESSION AP005939 BA000040
VERSION AP005939.1 GI:27349336
KEYWORDS
SOURCE
ORGANISM
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
1
REFERENCE
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,
Sasamoto, S., Watanabe, A., Idesawa, K., Iriiguchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)

TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)

JOURNAL
```

```
MEDLINE 22484998
PUBMED 12597275
REFERENCE 2
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,
Sasamoto, S., Watanabe, A., Idesawa, K., Iriiguchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)
TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL
MEDLINE 22485002
PUBMED 12597279
REFERENCE 3
AUTHORS
Kaneko, T.
Direct Submission
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp/rhizobase/,
URL: http://www.kazusa.or.jp/rhizobase/,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
Location/Qualifiers
1. 300900
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
103..756
/gene="blr1086"
103..756
/gene="blr1086"
/translation="MSKPTIEIRRAPVLGSSMAYRETGAQDPVVLHGNPTSSHI
WNILPLVSPVACHTAPDLIGFGQSGKPDIAFYDFHVRYLDAFIQRGVTSAYLVQ
DMGTALAFILAAARPDVFRGLAFMBETPMPTWODEFHTVEAEQDAEAAAFVRFK
RPFGEHMLLEAFVVERVLPGLVIRKLGDEMAEYRTEPPTPESRRVLIAPREL
IAGEADVDYDAALSAALASYPKLLFTGFGPGLVSPFEAFERFAASITRCALIRLG
AGHYLQEDHAALIGSVAGWAGTAEVVRPQAA"
1941..2372
/gene="blr1088"
1941..2372
/gene="blr1088"
/translation="MPVWNEDPAALDMNLSALYTAFLGNKTPADLALVFGNSMFC
AFAYDQRLVAGRALGDCAYLCDAVVRDCCQGLGRTIERLLAECRGRKII
LYAVPGKEGFEYRLGFRMTTAMAFDDQASAHQRGYLTAP"
2795..3850
/gene="blr1089"
MEDLINE 22484998
PUBMED 12597275
REFERENCE 2
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,
Sasamoto, S., Watanabe, A., Idesawa, K., Iriiguchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)
TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL
MEDLINE 22485002
PUBMED 12597279
REFERENCE 3
AUTHORS
Kaneko, T.
Direct Submission
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp/rhizobase/,
URL: http://www.kazusa.or.jp/rhizobase/,
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
Location/Qualifiers
1. 300900
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
103..756
/gene="blr1086"
103..756
/gene="blr1086"
/translation="MSKPTIEIRRAPVLGSSMAYRETGAQDPVVLHGNPTSSHI
WNILPLVSPVACHTAPDLIGFGQSGKPDIAFYDFHVRYLDAFIQRGVTSAYLVQ
DMGTALAFILAAARPDVFRGLAFMBETPMPTWODEFHTVEAEQDAEAAAFVRFK
RPFGEHMLLEAFVVERVLPGLVIRKLGDEMAEYRTEPPTPESRRVLIAPREL
IAGEADVDYDAALSAALASYPKLLFTGFGPGLVSPFEAFERFAASITRCALIRLG
AGHYLQEDHAALIGSVAGWAGTAEVVRPQAA"
1941..2372
/gene="blr1088"
1941..2372
/gene="blr1088"
/translation="MPVWNEDPAALDMNLSALYTAFLGNKTPADLALVFGNSMFC
AFAYDQRLVAGRALGDCAYLCDAVVRDCCQGLGRTIERLLAECRGRKII
LYAVPGKEGFEYRLGFRMTTAMAFDDQASAHQRGYLTAP"
2795..3850
/gene="blr1089"
```

CDS	2795..3850 /gene="blr1089" /note="ORF ID:blr1089" hypothetical protein /codon_start=1 /transl_table=11 /protein_id="BAC46354.1" /db_xref="GI:27349340" /translation="MLEAIRRAMTFLRQKVLHKLGVLSVAIVIGIACVLYVHLRGI DNEVEIAIKSTEPSQIAWALFVAAGYFTLYFDLFCAGRAIGHVPRINALAFT SYSIGHNVASVFTGAVRIYSYAGLNAIDVAKIFCLAGLTFWIGNAAVIGLGISY HPEAASIDQLPWLNLALMIIVGLGVVWVWTPQVRGVGWVTVVLPGLPFTLL QJAGIDIDGFCALAMVIVDEPNLGFVVAVIVFSATLLGFASHPGGLGVDAAM LVGLWMDRELLGMLLFRVLYLSPFVLSVILLTFREVIIGARSKRLQQAALKLDP GPAREAAVYRERSDGTG" 3910..5208 /gene="phoR" 3910..5208 /gene="phoR" /note="ORF ID:blr1090" /codon_start=1 /transl_table=11 /product="phosphate regulon, two-component sensor histidine kinase" /protein_id="BAC46355.1" /db_xref="GI:27349341" /translation="MAIDAFSSPTAQPSDRLHSTIIIAAALSVVVSIGLSVL HAVAFLCTAAALIPWRLHDTAASDDVRRINPVSAAVAAGMPDPVILLDRAG RVTHLAAQAALPALERNEIAQFALRSPBIIITALERSIATTEPRATYLDHVPVDRW MELIITPVPTTFFGGADKMLMTFHDQTPLRVEMRADFVANASHELPTPLAALSG FIDTLOQAKDDPKARERFIIMHNOATRMARLIDLLSLRVELSAHVRPDTIVDLL PIRQVADGLEPLARERQVEVEIHLPEPPVMIAGDRELLRLPENLLENALYKAGSGG RVIVLSTGAATDGTQIRVWRVDFGEGIAPEHLPLRTERFYRVVDGDSRLQGTGVLG LSLVKHILNRHGRLELLIESVPKQATFACFPQIRPPVTV" 5341..6366 /gene="pstS" 5341..6366 /gene="pstS" /note="ORF ID:blr1091" /codon_start=1 /transl_table=11 /product="ABC transporter phosphate-binding protein" /protein_id="BAC46356.1" /db_xref="GI:27349342" /translation="METSMPNLTIVAGLVAATTAADYAGADYAGATPPPIYSKW ADAYKKTGSLNTQSIGGGIKQIQAKTVTFGASDAPLKAQLEKDKGLVQWFWMG AIVPVNIEGVKPGEMVDFGETLASIYLGKITKWDAAIKKLNPLKLPDALITVRR SDGSGTTFNFNYLSKASADWKSXGEGTAVWPGVGAKEGVSGNIGTKNSIGY VEYAYAKNKLTYTGLYNKACKPVQPTVEAFQAAASNADWAKAPGYVILTDQGEKS WPTTAATFILMKDQATKASQEAALKPFRWAFKNGKAAEELDYIPMEGVVTLIEKT WAAEIKS" 6487..7485 /gene="pstC" 6487..7485 /gene="pstC" /note="ORF ID:blr1092" /codon_start=1 /transl_table=11 /product="phosphate ABC transporter permease protein" /protein_id="BAC46357.1" /db_xref="GI:27349343" /translation="MAEMAVQSDVIDDAGPYDRAKALSAFKLGDVTFYVITLSAISV LLILGGLIVSLIVGAFPAIKYRFLFTQWAPADPPVLGALGPMYGTVTSFIAM LIAIPVGLGIAITFELCPQWLRFIMAGIELLAGIYFIYGMWGFVGLGPIPLANTFQ PFMIKIFDGVVLGIAFAGPSYLSLFAALILAIMLVLPFTISIVDFKTVFVKLKE AAGVGTCTWEVRSWVITPYRVIGVIGVLMALGALGETMAVTFIIGNSPFISSIF APGTTISAIASEFAESDGLHQSGILILGLLFLVLTFTFVLAARLMLRLEKKAGN" 7490..8335 /gene="pstA" 7490..8335 /gene="pstA" /note="ORF ID:blr1093" /codon_start=1	gene
CDS	8332..9153 /gene="pstB" 8332..9153 /gene="pstB" /note="ORF ID:blr1094" /codon_start=1 /transl_table=11 /product="phosphate ABC transporter ATP-binding protein" /protein_id="BAC46359.1" /db_xref="GI:27349345" /translation="MSELSVMSAAGLPQAPLLPEAPKVTVRNLFYGEHHLAKN INLTGTRVTAFIGPSGGCKSTLLIFNRMYDLYPQORATQGLMDLTNLDKLDL NLIRARVGUVFOKPTPPMTIYENIAFGIRLYEKISKEMDDRVEKALRGALNWEYK DKLNASGLSLSGQQORQLCIARTVAVRPSVILFDEPCSLDPISTAKVELIOELSEN YTIIVTHNMQQARVSDKIAFMYLGEILIEFDTSKIFTSPSRRTQDYITGREG" 9165..9881 /gene="phoU" 9165..9881 /gene="phoU" /note="ORF ID:blr1095" /codon_start=1 /transl_table=11 /product="phosphate transport system regulatory protein" /protein_id="BAC46360.1" /db_xref="GI:27349346" /translation="MGSEHTAKAFDPLQLTRELVAEMGGIAERMIVDSVDALIRRDV	gene
CDS	9165..9881 /gene="phoU" 9165..9881 /gene="phoU" /note="ORF ID:blr1095" /codon_start=1 /transl_table=11 /product="phosphate transport system regulatory protein" /protein_id="BAC46360.1" /db_xref="GI:27349346" /translation="MGSEHTAKAFDPLQLTRELVAEMGGIAERMIVDSVDALIRRDV	CDS
gene	9165..9881 /gene="phoU" 9165..9881 /gene="phoU" /note="ORF ID:blr1095" /codon_start=1 /transl_table=11 /product="phosphate transport system regulatory protein" /protein_id="BAC46360.1" /db_xref="GI:27349346" /translation="MGSEHTAKAFDPLQLTRELVAEMGGIAERMIVDSVDALIRRDV	gene
Query Match	4.0%; Score 20; DB 1; Length 300900;	
Best Local Similarity	100.0%; Pred. No. 65;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	13 TCGGGCACGACCGGGCGGCC 32	
Db	199354 TCGGGCACGACCGGGCGGCC 199373	
RESULT 39		
AP005034/c		
LOCUS	AP005034 302300 bp DNA linear BCT 10-MAY-2003	
DEFINITION	Streptomyces avermitilis genomic DNA, complete genome, section 14/30.	
ACCESSION	AP005034 BA000030	
VERSION	AP005034.1 GI:29606779	
KEYWORDS	Streptomyces avermitilis MA-4680	
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetes.	
ORGANISM	1 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osone,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites	
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)	
AUTHORS	21477403 PUBLISHED 11572948	
TITLE	Streptomyces avermitilis MA-4680	
JOURNAL	Streptomyces avermitilis MA-4680	
MEDLINE	21477403 PUBLISHED 11572948	
REFERENCE	Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis	



JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306  
12692562  
3 (bases 1 to 302300)  
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,  
Sakaki, Y., and Hattori, M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center,  
National Institute of Technology and Evaluation, Biotechnology  
Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/,  
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun  
Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi  
Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi  
Osonoe(\*4), Norihiro Kushida(\*4), Hisashi Kikuchi(\*4), Tadayoshi  
Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7)  
and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda  
and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermittilis.ls.kitasato-u.ac.jp.

FEATURES  
source

Location/Qualifiers  
1. .302300  
/organism="Streptomyces avermiltis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/notes="This strain is also named as strain: ATCC 31267,  
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces  
avermectinii"  
/gene="fdxD"  
/complement(142. .462)  
/gene="fdxD"  
/complement(142. .462)  
/note="SAV3129"  
/codon\_start=1  
/transl\_table=11  
/product="putative ferredoxin"  
/protein\_id="BAC70840.1"  
/db\_xref="GI:29606780"  
/translation="MYVIAEPCVDVKAKIECPVDDCIYEGSRSLYIHPDRCVDCG  
ACPEVCPVEAIFYDDTPPEWKDYKANVEFFDELSPGSGSKGLIERDHPFIALP  
PNQ"  
603. .1976  
/note="SAV3130"  
/codon\_start=1  
/transl\_table=11  
/product="putative acetyltransferase"  
/protein\_id="BAC70841.1"  
/db\_xref="GI:29606781"  
/translation="MEISAGQLEVRITADVGKRVSVRWNNEDVAGGEKFTDTGVYL  
TSWDNGVLLITRGGESVRIAESLVAGKVPAPARRRGPAASEYELARARAWQP  
VERRIGDWGLRASAGTRTRANSVLPVGEPLGLDEALTIVRDWYAGRELPAVLOTAT  
GAEGTQLLCALDEARGTWREVTAEIWLGIAPLADREVAEVLSEADWLRYQ  
RKGLNEVALVKGSPSWFATVPCGAAGRAAPAAIGRCVGDWRWAGFAAEVDPAQRR  
GLATVMTLARRALEGSAAMQLQVETDNAAALYAGMGFATHAYHYRWASED  
VGLGGRERGAPEAGVSGEAGVDEAGVDEAGVDEAGVDEAGVDEAGVDEAGVDEAGV  
GALGSGAPRISTPDDVSSAAGESDVECGSGSAGSVGGPGSGGSGGSDAVDG  
VGTRSV"  
1973. .2845  
/note="SAV3131"  
/codon\_start=1  
/transl\_table=11

CDS

product="hypothetical protein"  
/protein\_id="BAC70842.1"  
/db\_xref="GI:29606782"  
/translation="MNGHEAAMPHPPEPERAAELRRRFADEARSRPDLISALCLLV  
GAADGALDEGIDAQIDDLRAGQVFFRGGPRPWALAEILLGGCGRGTTPADY  
QRLESSLHEVIRRRRGLPILLSVMEVARRAGAPVGVVALPGHFVVGGPBPQOVL  
ADPFGGRVLTGGDAELLVAGATGAPLDPMSLTADPLDVVLRLINIRAWAARPER  
SDVALMVDLSLLPSHPARLYDRYQQLVERGDFLTGAMELDAYAEVVEADGSAAD  
RIRHQAAARSMLN"  
complement(2864. .3478)  
/note="SAV3132"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system response  
regulator"  
/protein\_id="BAC70843.1"  
/db\_xref="GI:29606783"  
/translation="MSRIKVLLEADQSMVREALAALGLEPDIIEVVAQVAGDRVLA  
AAREHAVDALLDIEMPGATGIEAAQNRRELPAVKLVLTTFGRPGVLSRAMEAGLA  
AFLVKADPAQALAEARVKLAGERVIDPTLAAALAGGANPLUTDREREVRANAAGAT  
NAELATALHLSQGTVRNYLSTAIQKLAVRNETEAVRTARDKGWL"  
complement(3489. .4646)  
/note="SAV3133"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system sensor kinase"  
/protein\_id="BAC70844.1"  
/db\_xref="GI:29606784"  
/translation="MRTGSPENVREALRKAMWIVLWVFLSSPVHDLASGHTTPAAT  
AAGWGLAFAFGVIGYITLVFRNMGRLNGPAVALVAGLVAVLCLSGPMLGLFV  
YVSAAGTTFPLRVAYITPLTAVAMLLVGLHGGQDARNIVLLVILGFAMTGVRL  
VRTVELKARATQAALANEERLRLDLHLGHSUSLTIKLSLGRLLADLPHEK  
AAQVADIEQVSQALVDREAVTGVRRCLAGELAGMQVALTAAGVTDADLPABDLD  
GVPEESGALAWALREAVTNVVRHSGARRCVVELVHROTLDGFLVLELSDVEDGSGSG  
STPGNLGTLTERLEKAGSLDAGVRRGRFLVARVPMASARGPVCTAGTP"  
complement(4697. .5455)  
/note="SAV3134"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter integral membrane  
protein"  
/protein\_id="BAC70845.1"  
/db\_xref="GI:29606785"  
/translation="MTNIFASWGLIKLEIVRALNRKFLFPFVIPSALFLIAGSAD  
STTKVEGTGLSLPTFFVMSAGFALTAVLMGNSRIKAKERESGWQRLLTLPGRG  
VYLKATAAVALSLPSILIVFAAAVKNLMDAQWLALTGTITWGSILVFAALGVAI  
GYLATGDAVRPITMIVYFGLSMGLGMLPPTTTPDWLQSTAKWLPTHAVAAALQAIEL  
GNAPQAKDIALAVSFVLEPAGGAWLYRKDTLKA"  
complement(5452. .6360)  
/note="SAV3135"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter ATP-binding protein"  
/protein\_id="BAC70846.1"  
/db\_xref="GI:29606786"  
/translation="MTTGSVAFDQVTKYVDVRAVDGLTLALHPGTVALLPGNGAG  
KSTLLDLLLKHADSGTVLFGSPREAIAGRVAGMLQSGGMLDEVTVEGLVKAC  
DLHKPFRVTDVLARAGVSIADRVKNLKGQEQVRVFALATAGTDLTLDDEPTFG  
MDVTRHAFWATREQADQRTVLFATHYLEADAIADRVILVHGRLLADGTAIEIK  
AKAGARVSFDLSPIDEALRALPELTISVDISGHTIRIQSTDADA\*VHALYGLGVFP  
RNLEVAGLGLGEQARFALTAABESKSK"  
complement(6662. .8941)  
/note="SAV3136"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC70847.1"  
/db\_xref="GI:29606787"  
/translation="MSRETDSSAGPNRGGAAYPGTTPYGTPTMAESDGVDAGRPAA  
KSDERKTETTLTTRINIPGRPIPPVNMRTFVDAEADAPAAAPAPGPVSAIS  
TYVPAASAPPAPGAPAAAEETSDWFAPRKSGAPKPGCGSTNGAGLPTGS  
GAQAPCGAGGGGPRPGAGGPRPGAGGPRPGAGGPRPGAGGPRPGAGGPRPGAGG  
QGQGGGTNGAGLPGATGAGVPAGHGGGTSGSFVSEAVATTSFPFLPVVGSNDGAEP

CDS

product="hypothetical protein"  
/protein\_id="BAC70842.1"  
/db\_xref="GI:29606782"  
/translation="MNGHEAAMPHPPEPERAAELRRRFADEARSRPDLISALCLLV  
GAADGALDEGIDAQIDDLRAGQVFFRGGPRPWALAEILLGGCGRGTTPADY  
QRLESSLHEVIRRRRGLPILLSVMEVARRAGAPVGVVALPGHFVVGGPBPQOVL  
ADPFGGRVLTGGDAELLVAGATGAPLDPMSLTADPLDVVLRLINIRAWAARPER  
SDVALMVDLSLLPSHPARLYDRYQQLVERGDFLTGAMELDAYAEVVEADGSAAD  
RIRHQAAARSMLN"  
complement(2864. .3478)  
/note="SAV3132"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system response  
regulator"  
/protein\_id="BAC70843.1"  
/db\_xref="GI:29606783"  
/translation="MSRIKVLLEADQSMVREALAALGLEPDIIEVVAQVAGDRVLA  
AAREHAVDALLDIEMPGATGIEAAQNRRELPAVKLVLTTFGRPGVLSRAMEAGLA  
AFLVKADPAQALAEARVKLAGERVIDPTLAAALAGGANPLUTDREREVRANAAGAT  
NAELATALHLSQGTVRNYLSTAIQKLAVRNETEAVRTARDKGWL"  
complement(3489. .4646)  
/note="SAV3133"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system sensor kinase"  
/protein\_id="BAC70844.1"  
/db\_xref="GI:29606784"  
/translation="MRTGSPENVREALRKAMWIVLWVFLSSPVHDLASGHTTPAAT  
AAGWGLAFAFGVIGYITLVFRNMGRLNGPAVALVAGLVAVLCLSGPMLGLFV  
YVSAAGTTFPLRVAYITPLTAVAMLLVGLHGGQDARNIVLLVILGFAMTGVRL  
VRTVELKARATQAALANEERLRLDLHLGHSUSLTIKLSLGRLLADLPHEK  
AAQVADIEQVSQALVDREAVTGVRRCLAGELAGMQVALTAAGVTDADLPABDLD  
GVPEESGALAWALREAVTNVVRHSGARRCVVELVHROTLDGFLVLELSDVEDGSGSG  
STPGNLGTLTERLEKAGSLDAGVRRGRFLVARVPMASARGPVCTAGTP"  
complement(4697. .5455)  
/note="SAV3134"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter integral membrane  
protein"  
/protein\_id="BAC70845.1"  
/db\_xref="GI:29606785"  
/translation="MTNIFASWGLIKLEIVRALNRKFLFPFVIPSALFLIAGSAD  
STTKVEGTGLSLPTFFVMSAGFALTAVLMGNSRIKAKERESGWQRLLTLPGRG  
VYLKATAAVALSLPSILIVFAAAVKNLMDAQWLALTGTITWGSILVFAALGVAI  
GYLATGDAVRPITMIVYFGLSMGLGMLPPTTTPDWLQSTAKWLPTHAVAAALQAIEL  
GNAPQAKDIALAVSFVLEPAGGAWLYRKDTLKA"  
complement(5452. .6360)  
/note="SAV3135"  
/codon\_start=1  
/transl\_table=11  
/product="putative ABC transporter ATP-binding protein"  
/protein\_id="BAC70846.1"  
/db\_xref="GI:29606786"  
/translation="MTTGSVAFDQVTKYVDVRAVDGLTLALHPGTVALLPGNGAG  
KSTLLDLLLKHADSGTVLFGSPREAIAGRVAGMLQSGGMLDEVTVEGLVKAC  
DLHKPFRVTDVLARAGVSIADRVKNLKGQEQVRVFALATAGTDLTLDDEPTFG  
MDVTRHAFWATREQADQRTVLFATHYLEADAIADRVILVHGRLLADGTAIEIK  
AKAGARVSFDLSPIDEALRALPELTISVDISGHTIRIQSTDADA\*VHALYGLGVFP  
RNLEVAGLGLGEQARFALTAABESKSK"  
complement(6662. .8941)  
/note="SAV3136"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC70847.1"  
/db\_xref="GI:29606787"  
/translation="MSRETDSSAGPNRGGAAYPGTTPYGTPTMAESDGVDAGRPAA  
KSDERKTETTLTTRINIPGRPIPPVNMRTFVDAEADAPAAAPAPGPVSAIS  
TYVPAASAPPAPGAPAAAEETSDWFAPRKSGAPKPGCGSTNGAGLPTGS  
GAQAPCGAGGGGPRPGAGGPRPGAGGPRPGAGGPRPGAGGPRPGAGGPRPGAGG  
QGQGGGTNGAGLPGATGAGVPAGHGGGTSGSFVSEAVATTSFPFLPVVGSNDGAEP

CDS

Thu Jun 17 09:02:21 2004

```

RDNLPIFSSENDRRPGSGTGGVTDGPAVPPRGCGGCGLGPAGCMSSPGGGLTAPG
TGPRPGVGPAGTGPAGCGSGGLSDDTAILTPQRPAPDPDPGCGGDNVSGH
TVTSGIPVPGQSGNSPFGGAGHTDGPVHTPKLPAPSEPKTATSSPKAKKGRSKL
VLLGGVFLGGVTVGAGLNMHSDVPKGTTLVGDIGGTRDEAVKLDLDDALGRTG
KALKSLVDGDTVSLPDQAGLQDSKATVSAANSYNEVSVIGSLFGQQRVIEPMP
VDEEKLQALERAAGAGSASGDTIKFTSGKAVAVYKAGKGLDVAAQATKAVEQART
QVSESDTSPVKVPTTTKPTVSNAEVDRMKKFAKPMASANTVQTDAAHAIAFSPON
SLWKFLGVAVANGKLVYADRAALKSLYGNITFDGLVLTGRTGKTKTAVTVEDVISALRP
ALLTTNGAARTAAVIDTDPG"
complement(9107..9526)
/notes="SAV3137"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="BAC70848.1"
/db_xref="GI:29606788"
/translation="MSTGRQSQALAQPVTLPSAGRAAAAYLGLLGLVVLGALGVQ
AAWPGGLLALAGAAGLHLGGSVVIGSRAGAVAPAGMVAVVLLTASRPGDFLFG
AGVSVYLLGGMAVAVMCATLAQGRPGGDRVLGK"
complement(9526..10410)
/notes="SAV3138"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70849.1"
/db_xref="GI:29606789"
/translation="MKLDPRRLHLVHAHPDSEINNGATWARYAEGAQVTLVTCVL
GECEVIPPALAHLTDPDLDDALGPRVYVGEALAAKMLGVTDRFLGGGRVDRSGMVG
GEQNRHAGAFWAADLDERAGHLVAVREVPRQVLTYPDGGYGHDPHQHRAVAVRA

```

```

Query Match 4.0%; Score 20; DB 1; Length 302300;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 ACCGGGCGGCCAAGGCGT 41
Db 30104 ACCGGGCGGCCAAGGCGT 30085
|||||

```

```

RESULT 40
AX756416/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1
Zhu, T., Cheng, W., Briggs, S., Cooper, B., Goff, S.A., Moughamer, T.,
Glazebrook, J., Katagiri, F., Kreis, J., Provart, N. and Ricke, D.
Identification and characterization of plant genes
Patent: WO 0300905-A 1155 03-JAN-2003;
Syngenta Participations AG (CH)
LOCATION/Qualifiers
1..558
/organism="Zea mays"
/mol_type="unassigned DNA"
/db_xref="taxon:4577"

```

```

Query Match 3.8%; Score 19; DB 6; Length 558;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 271 GCCGGCGTCCGACCACT 289
Db 395 GCCGGCGTCCGACCACT 377
|||||

```

```

RESULT 42
AF481794/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Streptomyces griseus
Streptomyces griseus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE
1 (bases 1 to 716)
Clark, D.J., Sigmund, J. and Anderson, A.S.
Detection of Bacterial 3-Hydroxy-3-Methylglutaryl Coenzyme A
Reductases from Actinomycetes
Unpublished
2 (bases 1 to 716)
Anderson, A.S.
Direct Submission
JOURNAL
REFERENCE
AUTHORS
TITLE

```

```

RESULT 41
AF267643/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 579)
Guo, S., Zhao, W., Li, Q. and Zhang, X.
Mxw gene cloning and expression analysis in Zea mays
Unpublished
2 (bases 1 to 579)
Guo, S., Zhao, W., Li, Q. and Zhang, X.
Direct Submission
Submitted (12-MAY-2000) College of Life Sciences, Shandong
Agricultural University, Daizhong Street 61, Taian, Shandong
271018, China
LOCATION/Qualifiers
1..579
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Huangzao 4"
/db_xref="taxon:4577"
/tissue_type="endosperm"
<1..>579
/genes="mxw"
/notes="synonym: waxy"
<1..>579
/genes="mxw"
/codon_start=1
/product="starch synthase GBSSI"
/protein_id="AAF79205.1"
/db_xref="GI:8778105"
/translation="VSTPVKAKLKEALQAEVGLPVDNRNIPAWAFIQLBEQKGPD
VNAAIIPQLMENVEDVQIVLLGTGKKKFERMLMSAEKEPKGVRAVVKNAALAHIM
AGADVLAVTSRFPQGLIQLQGMRYGTPCASCATGGLVDTIIEGKTGFENGRLSVDGN
VVEPADVKVAVTTLQRAIKVGTTPAYEVNRNC"
Query Match 3.8%; Score 19; DB 8; Length 579;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 271 GCCGGCGTCCGACCACT 289
Db 554 GCCGGCGTCCGACCACT 536
|||||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:31:50 ; Search time 246 Seconds

(without alignments)  
8651.818 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501  
Sequence: 1 tcgtcttcgtgctgcgcgt.....cgctgtgctgcgggtcggtg 501

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq\_26Jan04.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002s.\*
- 7: Geneseqn2003as.\*
- 8: Geneseqn2003bs.\*
- 9: Geneseqn2003cs.\*
- 10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	501	100.0	53789	2	Aav21187 Amycolato
2	36	7.2	20394	4	Aaf24892 Pimaricin
3	36	7.2	43280	2	Aat80413 Ty lactone
C 4	33	6.6	103599	4	Abx04971 S. cinna
5	32	6.4	29879	2	Aaq46806 eryA regi
C 6	29	5.8	77536	3	Aal14651 Nucleotid
7	29	5.8	113193	7	Aad54645 Streptom
8	28	5.6	9222	7	Aal61171 Actinosyn
C 9	28	5.6	82746	7	Aal61224 Actinosyn
10	27	5.4	665	4	Aaf89965 Partial n
11	27	5.4	15872	2	Aat68715 Streptom
12	27	5.4	15872	3	Aaz87283 S. venez
13	26	5.2	125401	4	Aad17186 Streptom
14	25	5.0	9975	7	Aal61173 Actinosyn
15	25	5.0	50937	3	AAAO9469 Streptoco
16	24	4.8	640	4	AAa89130 Polyketid
17	24	4.8	3978	3	RAC55785 Type I po
18	24	4.8	3978	9	ADE10204 S. lavend
19	24	4.8	12249	3	RAC55840 Complete
20	24	4.8	12249	9	ADE10259 S. lavend
21	24	4.8	18331	3	RAC55857 Complete
22	24	4.8	18331	9	ADE10266 S. lavend
23	24	4.8	103599	4	ABX04971 S. cinna

Aa14666 Nucleotid  
Aa14665 Nucleotid  
Aa14668 Nucleotid  
Aa14667 Nucleotid  
Aa161172 Actinosyn  
Aa14669 Nucleotid  
Aa21186 Amycolato  
Aaf88335 S. spinos  
Aad80415 Hybrid sr  
Aad26981 Sorangium  
Aad17185 Streptom  
Aa92301 S. avermi  
Aah79277 Streptom  
Aa92302 S. avermi  
Aah79278 Streptom  
Aa17367 DNA seque  
Aat80414 Platenoli  
Aat78508 Platenoli  
Aaf88312 S. spinos  
Aaf88316 S. spinos  
Aad17184 Streptom  
Aad26995 Sorangium

ALIGNMENTS

RESULT 1

AAV21187  
ID AAV21187 standard; DNA; 53789 BP.

XX

AC AAV21187;

XX 24-JUL-1998 (first entry)

DT

XX

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XW polyketide synthase; actinomycete; ansamycin; ds.

XX

OS Amycolatopsis mediterranei.

XX

Key Location/Qualifiers

FH 1825..15543

CDS /tag= a

FT /label= ORF\_A

FT /product= "polyketide synthase"

FT 15550..30759

FT /tag= b

FT /label= ORF\_B

FT /product= "polyketide synthase"

FT 30895..36060

FT /tag= c

FT /label= ORF\_C

FT /product= "polyketide synthase"

FT 36259..41325

FT /tag= d

FT /label= ORF\_D

FT /product= "polyketide synthase"

FT 41373..51614

FT /tag= e

FT /label= ORF\_E

FT /product= "polyketide synthase"

FT 51713..52993

FT /tag= f

FT /label= ORF\_F

FT /product= "polyketide synthase"

FN WO9807868-A1.

XX 26-FEB-1998.

PD

XX 18-AUG-1997;

PF 97WO-EP004495.



XX Ty lactone synthase gene cluster.

DE Ty lactone synthase gene cluster; tylG gene; multifunctional protein;

XX polyketide; ty lactone synthesis; antibiotic; tylosin; ss.

XX Streptomyces fradiae.

XX Key Location/Qualifiers

FT 816..14243

FT /tag= a

FT /transl\_except= (pos: 816..818, aa: Met)

FT /note= "ORF1 encodes protein shown in AAW22601"

FT 14351..19945

FT /tag= b

FT /transl\_except= (pos: 14351..14353, aa: Met)

FT /note= "ORF2 encodes protein shown in AAW22602"

FT 20010..31199

FT /tag= c

FT /transl\_except= (pos: 20010..20012, aa: Met)

FT /note= "ORF3 encodes protein shown in AAW22603"

FT 31232..36067

FT /tag= d

FT /note= "ORF4 encodes protein shown in AAW22604"

FT 36249..41774

FT /tag= e

FT /note= "ORF5 encodes protein shown in AAW22605"

FT EP791655-A2.

XX 27-AUG-1997.

XX 19-FEB-1997; 97EP-00301056.

XX 22-FEB-1996; 96US-0012078P.

XX (ELIL ) LILLY & CO ELI.

XX Dehoff BS, Kuhstoss SA, Rostek PR, Sutton KL;

XX WPI; 1997-418046/39.

XX P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.

XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for

XX production of tylosin-related polyketide compounds.

XX Claim 2; Page 8-66; 220pp; English.

XX This sequence represents the ty lactone synthase gene cluster of the

XX invention. This sequence is also referred to as the tylG gene, and was

XX isolated from Streptomyces fradiae. This sequence encodes multifunctional

XX proteins which direct the synthesis of the polyketide ty lactone, isolated

XX from Streptomyces fradiae. Ty lactone is the basic building block of the

XX antibiotic tylosin. The DNA sequence can be modified so as to alter the

XX type of carboxylic acids incorporated, the number of carboxylic acids

XX incorporated and/or the post-condensation reactions performed, thereby

XX resulting in novel tylosin-related polyketides

XX Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 U; 0 Other;

XX

Query Match 7.2%; Score 36; DB 2; Length 43280;

Best Local Similarity 100.0%; Pred. No. 1.4e-06;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGGCCCCCGACGGCGCTGCAAGCGTTCGCCGAC 168

Db 15074 CTGGCCCCCGACGGCGCTGCAAGCGTTCGCCGAC 15109

RESULT 4

ABX04971/c

ID ABX04971 standard; DNA; 103599 BP.

XX

AC ABX04971;

XX 16-JAN-2003 (first entry)

XX S. cinnamomensis monensin type I polyketide synthase gene cluster.

DE Monensin; gene; cluster; polyketide synthase; antibiotic; ds;

XX antihelminthic; insecticide; immunosuppressant; antifungal;

KW antibacterial; polyether; mon BI; mon BI; mon CI; mon CII; mon H;

XX mon RI; mon RII; mon T; mon AIX; mon AX.

XX Streptomyces cinnamomensis.

OS WC200168867-A1.

PN 20-SEP-2001.

PD 30-MAY-2000; 2000WO-GB002072.

XX 28-MAY-1999; 99GB-00012563.

XX (BIOT-) BIOTICA TECHNOLOGY LTD.

XX Leadlay PF, Staunton J, Oliynyk M;

XX WPI; 2001-611393/70.

DR P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859,

DR ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866,

DR ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873,

DR ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880,

DR ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887.

XX New DNA sequence encoding polyketide synthase, useful for the production

XX of polyketides such as antibiotic monensin.

XX Claim 1; Page 116-195; 212pp; English.

XX The invention relates to a DNA sequence which is a fully defined sequence

XX of 103551 base pairs appearing as ABX04971, or its variant, that it is

XX not a sequence encoding all or part amino acids 1-920 encoded by mon AI

XX as given in the specification. The DNA is the S. cinnamomensis polyketide

XX antibiotic monensin biosynthetic gene cluster. Also included are a

XX recombinant cloning or expression vector comprising the gene cluster, a

XX transformant host cell which has been transformed to contain the gene

XX cluster (and is capable of expressing a corresponding polypeptide), a

XX hybridization probe derived from the gene cluster (for identification and

XX isolation of the same or analogous gene cluster, e.g. one which binds

XX specifically to a region of the monensin gene cluster selected from mon

XX BI, mon BII, mon CII, mon H, mon RI, mon RII, mon T, mon AIX and

XX mon AX), the use of the mon RI gene or variant and a monensin promoter to

XX control expression of a heterologous gene in Streptomyces cinnamomensis,

XX a polypeptide encoded by a portion of the monensin gene cluster

XX (preferably comprising mon BI, mon BII, mon CII or mon AX or their

XX mutants, alleles or variants), an epoxide enzyme encoded by mon CI, a

XX cyclase enzyme encoded by mon CII, producing S. cinnamomensis capable of

XX enhanced levels of production of monensin comprising engineering it to

XX overexpress the mon RI gene, S. cinnamomensis containing multiple copies

XX of the mon RI gene and/or its variants, expressing a gene heterologous to

XX S. cinnamomensis comprising transforming S. cinnamomensis with DNA

XX encoding a heterologous gene and expressing the gene under control of the

XX activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The

XX processes and materials (enzyme systems, nucleic acids and vectors) are

XX useful for preparing polyketides by recombinant synthesis. The

XX polyketides are useful as insecticides, antibiotics, antihelminthics,

XX antifungals, antibacterials or other pharmaceuticals. In particular, the

XX gene is useful for the production of monensin, an antibiotic polyether

XX polyketide. The present sequence represents the monensin gene cluster

XX Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;

XX

Query Match 6.6%; Score 33; DB 4; Length 103599;

Best Local Similarity 100.0%; Pred. No. 2.7e-05;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 381 GGAGTCGACGCGGTGGAGCGGCACGCGCACCGG 413  
|||||  
DB 83065 GGAGTCGACGCGGTGGAGCGGCACGCGCACCGG 83033  
|||||  
RESULT 5  
RAQ46806  
ID AAQ46806 standard; DNA; 29879 BP.  
XX  
AC AAQ46806;  
XX  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 22-DEC-1993 (first entry)  
XX  
DE eryA region of S. erythraea chromosome.  
XX  
KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;  
KW erythromycin; condensation; elongation; acyl chain growth;  
KW gene replacement; ss.  
XX  
OS Saccharopolyspora erythraea.  
XX  
FH Key Location/Qualifiers  
FT CDS 3..10199  
FT /tag= a  
FT /note= "ORF 1"  
FT 10218..20921  
FT /tag= b  
FT /note= "ORF 2"  
FT 20922..29879  
FT /tag= c  
FT /note= "ORF 3"  
XX  
PN WO9313663-A1.  
XX  
XX 22-JUL-1993.  
XX  
PF 17-JAN-1992; 92WO-US000427.  
XX  
PR 17-JAN-1992; 92WO-US000427.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
XX Katz L, Donadio S, Mcalpine JB;  
XX  
XX WPI; 1993-242804/30.  
XX P-PSDB; AAR44430, AAR44431, AAR44432.  
XX  
XX Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -  
PT by introducing altered biosynthetic gene-contg. DNA into microorganisms.  
XX  
XX Claim 27; Fig 2; 133pp; English.  
XX  
XX This sequence represents a fragment of the Saccharopolyspora erythraea  
CC genome, designated eryA. The polypeptides encoded by this region are  
CC involved in the biosynthesis of the polyketide segment of erythromycin.  
CC eryA is organised in modules and each module takes care of one  
CC condensation step. The precise succession of elongation steps is dictated  
CC by the genetic order of the modules. This fragment may be specifically  
CC altered such that novel polyketide molecules of desired structure are  
CC produced. Three types of alteration may be produced; those inactivating a  
CC single function in a module which does not arrest acyl chain growth;  
CC those inactivating a single function in a module which does affect chain  
CC growth; and those affecting an entire module. The mutations may be  
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN  
CC field.) (Updated on 27-AUG-2003 to correct OS field.)  
XX  
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;  
SQ

Query Match 6.4%; Score 32; DB 2; Length 29879;  
Best Local Similarity 100.0%; Pred. No. 8.9e-05;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 382 GACGTGACGCGGTGGAGCGGCACGCGCACCGG 413  
|||||  
DB 11205 GACGTGACGCGGTGGAGCGGCACGCGCACCGG 11236  
|||||  
RESULT 6  
AAA14651/c  
ID AAA14651 standard; DNA; 77536 BP.  
XX  
AC AAA14651;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.  
XX  
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;  
KW FK-506 binding protein; polyketide compound; transplant rejection;  
KW graft-versus-host disease; uveitis; alopecia universalis;  
KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
KW peripheral neuropathy; ss.  
XX  
OS Streptomyces hygroscopicus.  
XX  
FH Key Location/Qualifiers  
FT CDS complement(412..11836)  
FT /tag= a  
FT /note= "fkBW gene"  
FT complement(2020..3579)  
FT /tag= b  
FT /note= "fkBV gene"  
FT 3969..4496  
FT /tag= c  
FT /note= "fkBR2 gene"  
FT complement(4595..5488)  
FT /tag= d  
FT /note= "fkBR1 gene"  
FT 5601..6818  
FT /tag= e  
FT /note= "fkBE gene"  
FT 6808..8052  
FT /tag= f  
FT /note= "fkBF gene"  
FT 8156..8824  
FT /tag= g  
FT /note= "fkBG gene"  
FT complement(9122..9883)  
FT /tag= h  
FT /note= "fkBH gene"  
FT complement(9894..10994)  
FT /tag= i  
FT /note= "fkBI gene"  
FT complement(10987..11247)  
FT /tag= j  
FT /note= "fkBJ gene"  
FT complement(11244..12092)  
FT /tag= k  
FT /note= "fkBK gene"  
FT complement(12113..13150)  
FT /tag= l  
FT /note= "fkBL gene"  
FT complement(13212..23988)  
FT /tag= m  
FT /note= "fkBC gene"  
FT complement(13452..13662)  
FT /tag= n  
FT /note= "ACP6"  
FT complement(13761..14394)  
FT  
FT misc\_feature  
FT misc\_feature

FT / \*tag= o  
FT /note= "KR6"  
FT complement(14517. .15294)  
FT / \*tag= p  
FT /note= "ER6"  
FT complement(15438. .16587)  
FT / \*tag= q  
FT /note= "dehydratase domain (DH) 6"  
FT complement(15887. .17820)  
FT / \*tag= r  
FT /note= "acyltransferase domain (AT) 6"  
FT complement(17820. .19053)  
FT / \*tag= s  
FT /note= "KS6"  
FT complement(19116. .19326)  
FT / \*tag= t  
FT /note= "ACP5"  
FT complement(19464. .20097)  
FT / \*tag= u  
FT /note= "KS5"  
FT complement(20241. .21420)  
FT / \*tag= v  
FT /note= "DH5"  
FT complement(21420. .22653)  
FT / \*tag= w  
FT /note= "AT5"  
FT complement(22653. .23892)  
FT / \*tag= x  
FT /note= "KSS"  
FT complement(23992. .46573)  
FT / \*tag= y  
FT /note= "fkB gene"  
FT complement(24163. .24373)  
FT / \*tag= z  
FT /note= "ACP4"  
FT complement(24997. .26146)  
FT / \*tag= aa  
FT /note= "DH4 (inactive)"  
FT complement(26146. .27430)  
FT / \*tag= ab  
FT /note= "AT4"  
FT complement(27430. .28684)  
FT / \*tag= ac  
FT /note= "KS4"  
FT complement(28750. .28960)  
FT / \*tag= ad  
FT /note= "ACP3"  
FT complement(28750. .28960)  
FT / \*tag= ae  
FT /note= "DH2 (inactive)"  
FT complement(29092. .29740)  
FT / \*tag= af  
FT /note= "KR3"  
FT complement(29869. .31018)  
FT / \*tag= ag  
FT /note= "DH3 (inactive)"  
FT complement(31018. .32185)  
FT / \*tag= ah  
FT /note= "AT3"  
FT complement(32185. .33439)  
FT / \*tag= ai  
FT /note= "KS3"  
FT complement(33505. .33715)  
FT / \*tag= aj  
FT /note= "ACP2"  
FT complement(33823. .34480)  
FT / \*tag= ak  
FT /note= "KR2"  
FT complement(34606. .35749)  
FT / \*tag= al  
FT /note= "AT2"  
FT complement(35749. .37144)  
FT / \*tag= am  
FT /note= "AT2"

FT misc\_feature complement(37145. .38296)  
FT / \*tag= an  
FT /note= "KS2"  
FT complement(38371. .38581)  
FT / \*tag= ao  
FT /note= "ACP1"  
FT complement(38677. .39307)  
FT / \*tag= ap  
FT /note= "KR1"  
FT complement(39442. .40609)  
FT / \*tag= aq  
FT /note= "DH1"  
FT complement(40609. .41842)  
FT / \*tag= ar  
FT /note= "AT1"  
FT complement(41842. .43093)  
FT / \*tag= as  
FT /note= "KS of extender module 1 (KS1)"  
FT complement(43144. .43660)  
FT / \*tag= at  
FT /note= "ACP of loading domain"  
FT complement(43777. .44629)  
FT / \*tag= au  
FT /note= "ER of loading domain"  
FT complement(44974. .46573)  
FT / \*tag= av  
FT /note= "CoA ligase of loading domain"  
FT 46754. .47788  
FT / \*tag= aw  
FT /note= "fkB gene"  
FT 47785. .52272  
FT / \*tag= ax  
FT /note= "fkB gene"  
FT 52275. .71465  
FT / \*tag= ay  
FT /note= "fkB gene"  
FT 52362. .53576  
FT / \*tag= az  
FT /note= "KS7"  
FT 53577. .54716  
FT / \*tag= ba  
FT /note= "AT7"  
FT 54717. .55871  
FT / \*tag= bb  
FT /note= "DH7"  
FT 56019. .56819  
FT / \*tag= bc  
FT /note= "ER7"  
FT 56943. .57575  
FT / \*tag= bd  
FT /note= "KR7"  
FT 57710. .57920  
FT / \*tag= be  
FT /note= "ACP7"  
FT 57990. .59243  
FT / \*tag= bf  
FT /note= "KS8"  
FT 59244. .60398  
FT / \*tag= bg  
FT /note= "AT8"  
FT 60399. .61412  
FT / \*tag= bh  
FT /note= "DH8 (inactive)"  
FT

Query Match 5.8%; Score 29; DB 3; Length 77536;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 277 GCGGTCAACTCCGACGGCGGTCGAACGG 305  
|||  
Db 37516 GCGGTCAACTCCGACGGCGGTCGAACGG 37488

```
RESULT 7
AAD54645
ID AAD54645 standard; DNA; 113193 BP.
XX
AC AAD54645;
DT 26-JUN-2003 (first entry)
XX
DE Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KW Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX OS Streptomyces nodosus.
XX
XX Location/Qualifiers
FH complement(4. .1824)
FT /*tag= a
FT /product= "ABC transporter encoded by S. nodosus amphG
FT gene"
FT complement(1805. .3628)
FT /*tag= b
FT /product= "ABC transporter encoded by S. nodosus amphH
FT gene"
FT 3840. .4874
FT /*tag= c
FT /product= "GDP-mannose dehydratase encoded by S. nodosus
FT amphDIII gene"
FT 5042. .33574
FT /*tag= d
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT nodosus amphI gene"
FT 33584. .50518
FT /*tag= e
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 15, 16 and 17 encoded by S. nodosus
FT amphJ gene"
FT 50571. .56675
FT /*tag= f
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 18 and thioesterase encoded by S.
FT nodosus amphK gene"
FT 56829. .58019
FT /*tag= g
FT /product= "Cytochrome P450 encoded by S. nodosus amphL
FT gene"
FT 58139. .58648
FT /*tag= h
FT /product= "ORF1, hypothetical protein"
FT complement(58756. .59610)
FT /*tag= i
FT /product= "ORF2, hypothetical protein"
FT 59869. .61470
FT /*tag= j
FT /product= "ORF3, hypothetical protein"
FT /transl_except= (pos:59869. .59871, aa:Met)
FT complement(61798. .61995)
FT /*tag= k
FT /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT complement(62051. .63250)
FT /*tag= l
FT /product= "Cytochrome P450 encoded by S. nodosus amphN
FT gene"
FT complement(63250. .64308)
FT /*tag= m
FT /product= "NDP-sugar aminotransferase encoded by S.
FT nodosus amphDII gene"
FT complement(64324. .65775)
FT /*tag= n
FT /product= "Glycosyl transferase encoded by S. nodosus
FT amphDI gene"
FT /transl_except= (pos:65773. .65775, aa:Met)
FT 66081. .70319
FT CDS
```

```
FT /*tag= o
FT /product= "Polyketide synthase multienzyme housing
FT loading module encoded by S. nodosus amphA gene"
FT 70366. .79938
FT /*tag= p
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 1 and 2 encoded by S. nodosus by amphB
FT gene"
FT 79956. .112709
FT /*tag= q
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT nodosus by amphC gene"
FT
XX
XX WO200297082-A2.
XX
XX 05-DEC-2002.
XX
XX 27-MAY-2002; 2002WO-IE000071.
XX
XX 31-MAY-2001; 2001IE-00000527.
XX (UVDU-) UNIV COLLEGE DUBLIN.
XX
XX Caffrey JP;
XX
XX WPI; 2003-201271/19.
XX P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX AAE36129, AAE36130, AAE36131, AAE36132.
XX
XX Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX for preparing amphotericin derivative or analog antibiotic agent with
XX altered properties, in biosynthesis of polyketide other than
XX amphotericin.
XX
XX Claim 1; Page 52-114; 276pp; English.
XX
XX The invention relates to the gene cluster encoding the polypeptides
XX responsible for the biosynthesis of the polyene antibiotic amphotericin
XX (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX useful for preparing amphotericin derivatives or analogue antibiotic
XX agents with altered properties and in the biosynthesis of polyketides
XX other than amphotericin. amphDIII, amphDII or amphDI mutants are useful
XX for producing amphotericin derivatives glycosylated with alternative
XX sugars; amphDIII or amphDII gene sequences are useful in engineered
XX biosynthesis of perosaminyl-amphoteronolide B; amphDIII or amphDII and
XX amphN gene sequences are useful in the engineered biosynthesis of
XX perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDIII, amphDII
XX and amphDI gene sequences are useful for preparing polypeptides capable
XX of addition of mycosamine to a polyketide other than amphoteronolide A or
XX B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX The present sequence is S. nodosus amph biosynthetic gene cluster
XX
XX Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX
XX Query Match 5.8%; Score 29; DB 7; Length 113193;
XX Best Local Similarity 100.0%; Pred. No. 0.0016;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 175 GACGGCACCGCTGGTCCGAGGGCGTCGG 203
XX |||||
XX DB 51354 GACGGCACCGCTGGTCCGAGGGCGTCGG 51382
XX
XX RESULT 8
XX AAL61171
XX ID AAL61171 standard; DNA; 9222 BP.
XX
XX AAL61171;
XX
XX 22-SEP-2003 (first entry)
XX
```





CC nucleic acids. The nucleic acids are sources for sequences that encode  
 CC either operons involved in a metabolic pathway (specifically polyketide  
 CC synthesis) or polypeptides, particularly for production of therapeutic or  
 CC agricultural compounds, especially polyketide antibiotics. AAF89964-78  
 CC encode partial type I polyketide synthases, and were isolated using the  
 CC method of the invention

XX SQ Sequence 665 BP; 80 A; 247 C; 244 G; 94 T; 0 U; 0 Other;  
 Query Match 5.4%; Score 27; DB 4; Length 665;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 CCGACGGCGCTGCGAGCGCTCGCG 166  
 Db 389 CCGACGGCGCTGCGAGCGCTCGCG 415

RESULT 11  
 AAT68715  
 ID AAT68715 standard; DNA; 15872 BP.  
 XX AC AAT68715;  
 XX DT 01-SEP-1997 (first entry)  
 DE Streptomyces venezuelae polyketide synthase vep ORF1.  
 KW polyketide synthase; polyhydroxyalkanoate monomer synthase;  
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;  
 KW metabolic engineering; ss.  
 XX OS Streptomyces venezuelae.  
 XX FH Key Location/Qualifiers  
 XX CDS 20..13912  
 XX FT /\*tag= a  
 XX FT CDS 14056..14136  
 XX FT /\*tag= b  
 XX FT CDS 14148..15827  
 XX FT /\*tag= c  
 XX PN W09722711-A1.  
 XX PD 26-JUN-1997.  
 XX PF 18-DEC-1996; 96WO-US020119.  
 XX PR 19-DEC-1995; 95US-0008847P.  
 XX PA (MINU ) UNIV MINNESOTA.  
 XX PI Sherman DH, Williams MD, Xue Y;  
 XX WPI; 1997-341701/31.  
 XX DR P-PSDB; AAW19629, AAW19630, AAW00918.  
 XX Expression cassettes for production of poly:hydroxy:alkanoate(s) -  
 PT provide wide range of biodegradable polymers for medical or industrial  
 PT use.  
 XX PS Claim 54; Fig 23; 91pp; English.

XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide  
 CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see also  
 CC AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5' loading  
 CC module and a 3' end domain. Each of the sequenced modules includes a keto  
 CC -ACP, an acyltransferase, a dehydratase, a keto- reductase and an acyl  
 CC carrier protein domain. The gene cluster was cloned using a heterologous  
 CC hybridisation strategy from a genomic DNA library. A novel expression  
 CC cassette encoding the first module from the vep gene cluster and module 7  
 CC from the Streptomyces tylop gene cluster has polyhydroxyalkanoate (PHA)  
 CC monomer synthase activity and can be used for PHA prodn. in host (esp.

CC insect) cells for use as a biodegradable polymer

XX SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 U; 0 Other;  
 Query Match 5.4%; Score 27; DB 2; Length 15872;  
 Best Local Similarity 100.0%; Pred. No. 0.016;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTCTTCGCGGCGCGG 131  
 Db 3721 GTTCGTCGAGTCTTCGCGGCGCGG 3747

RESULT 12  
 AAZ87283  
 ID AAZ87283 standard; DNA; 15872 BP.  
 XX AC AAZ87283;  
 XX DT 15-SEP-2003 (revised)  
 DT 05-JUN-2000 (first entry)  
 XX DE S. venezuelae vep ORF 1, SEQ ID NO:1.  
 XX KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolaemia; crop protection agent; ds.  
 XX OS Streptomyces venezuelae; ATCC15439.  
 XX FH Key Location/Qualifiers  
 XX CDS 20..13912  
 XX FT /\*tag= a  
 XX FT CDS /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"  
 XX FT 14056..14151  
 XX FT /\*tag= b  
 XX FT CDS /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"  
 XX FT 14167..15827  
 XX FT /\*tag= c  
 XX FT CDS /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"  
 XX PN W0200000620-A2.  
 XX PD 06-JAN-2000.  
 XX PF 25-JUN-1999; 99WO-US014398.  
 XX PR 26-JUN-1998; 98US-00105537.  
 XX PA (MINU ) UNIV MINNESOTA.  
 XX PI Sherman DH, Liu H, Xue Y, Zhao L;  
 XX WPI; 2000-160679/14.  
 XX DR P-PSDB; AAY77177, AAY77178, AAY77199.  
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
 PT synthesis of methymycin and pikromycin.  
 XX PS Example 3; Fig 23; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the erc gene cluster of Saccharopolyspora erythraea or  
 CC Streptomyces antibioticus. The invention also relates to a macrolide  
 CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desosamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the production

of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents a Streptomyces venezuelae ATCC 15439 DNA sequence, designated vep ORF 1 in the specification, which actually contains 3 open reading frames, which encode proteins AAY7177-Y7178 and AAY7179. The vep ORF 1 protein is defined in the specification as a PHA monomer synthase. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 U; 0 Other;

Query Match 5.4%; Score 27; DB 3; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTCTCGCGCAGCGGG 131

Db 3721 GTTCGTCGAGTCTCGCGCAGCGGG 3747

## RESULT 13

AA17186  
ID AAD17186 standard; DNA; 125401 BP.

XX AAD17186;

XX 29-NOV-2001 (first entry)

DE Streptomyces noursei nystatin PKS gene cluster DNA.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KW antifungal; antibiotic; ds.

XX Streptomyces noursei.

XX Key Location/Qualifiers

FT 6337..34771

FT CDS

FT /tag= a

FT /product= "NysI complete protein"

FT 34792..51099

FT /tag= b

FT /product= "NysJ protein"

FT 51155..57355

FT /tag= c

FT /product= "NysK protein"

FT 57503..58687

FT /tag= d

FT /product= "NysL protein"

FT complement(58786..58980)

FT /tag= e

FT /product= "NysM protein"

FT /note= "CDS does not include start codon"

FT complement(59045..60241)

FT /tag= f

FT /product= "NysN protein"

FT /note= "CDS does not include start codon"

FT complement(60238..61296)

FT /tag= g

FT /product= "NysD2 complete protein"

FT 120628..121308

FT /tag= h

FT /product= "NysR4 (long) protein"

XX

XX WO200159126-A2.

XX

XX 16-AUG-2001.

XX

XX 08-FEB-2001; 2001WO-GB000509.

XX

XX 08-FEB-2000; 2000GB-00002840.

XX

XX 10-APR-2000; 2000GB-00008786.

XX

XX 14-APR-2000; 2000GB-00009387.

XX (UNNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE/) DZIEGLEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAERVIK E.

XX (BRAU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX (VALL/) VALLA S.

XX

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

XX Valla S, Ellingren TE, Sletta H, Gulliksen O;

XX

XX WPI; 2001-557614/62.

XX

XX P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,

XX AAE10149, AAE10150.

XX

XX New nystatin polyketide synthase polynucleotides and polypeptides, useful

XX as antibiotics and antifungals.

XX

XX Claim 1; Page 188-254; 266pp; English.

XX

XX The present invention relates to the cloning and sequencing of the gene

XX cluster encoding a modular type I polyketide synthase (PKS) enzyme

XX involved in the biosynthesis of the macrolide antibiotic nystatin. The

XX nystatin PKS is useful as antifungal antibiotics. The present sequence is

XX a Streptomyces noursei nystatin PKS gene cluster DNA

XX

XX SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;

XX

XX Query Match 5.2%; Score 26; DB 4; Length 125401;

XX Best Local Similarity 100.0%; Pred. No. 0.035;

XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX QY 142 GACGGCCGCTGCAAGCGTTCGCCGA 167

XX

XX Db 51905 GACGGCCGCTGCAAGCGTTCGCCGA 51930

XX

XX RESULT 14

XX AAL61173

XX ID AAL61173 standard; DNA; 9975 BP.

XX

XX AAL61173;

XX

XX 22-SEP-2003 (first entry)

XX

XX Actinosynnema pretiosum polyketide synthase (PKS) gene #4.

XX

XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;

XX gene; ds.

XX

XX Actinosynnema pretiosum.

XX

XX WO2003045312-A2.

XX

XX 05-JUN-2003.

XX

XX 21-NOV-2002; 2002WO-US037547.

XX PF



```
CC field.)
XX Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 U; 0 Other;
SQ Best Local Similarity 100.0%; Score 25; DB 3; Length 50937;
Query Match 5.0%; Score 25; DB 3; Length 50937;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCGCTGCAAGGCGTTGCGCG 166
DB 9582 GACGGCGCTGCAAGGCGTTGCGCG 9606

RESULT 16
AA89130
ID AA89130 standard; DNA; 640 BP.
XX
XX AA89130;
AC
XX 19-MAR-2001 (first entry)
DT
DE Polyketide synthase derived sequence 34-183R1AS2.T7 DNA.
XX
XX Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
KW immunosuppressive; neuroimmunophilin; ds.
XX
XX Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 1..639
FT FT /*tag= a
FT PT /partial
FT ET /note= "includes in-frame stop codon at 613. .615"
XX
XX US6150513-A.
XX
XX 21-NOV-2000.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Wu X;
XX
XX WPI; 2001-049104/06.
XX P-PSDB; AAB19897.
XX
XX Novel nucleic acid encoding polyketide synthase useful for producing
XX recombinant polyketide synthase for making novel polyketides for use as
XX antibiotic, immunosuppressant or as neuroimmunophilin.
XX
XX Example; Col 21; 30pp; English.
XX
XX The present sequence is that of a polyketide synthase (PKS)-derived
XX coding sequence designated 34-183R1AS2.T7. The DNA was isolated from
XX cosmid 34-183 of an FK-520-producing strain of Streptomyces hygroscopicus
XX following restriction digestion. It does not comprise a full-length
XX coding sequence. The 213-amino acid translated sequence is given in
XX AAB19897. An isolated nucleic acid comprising a coding sequence for a PKS
XX enzyme identical to, or isolated from, a PKS coding sequence contained
XX within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is
XX claimed. Such nucleic acids, when used alone or in combination with other
XX PKS domain coding sequences, are useful in the construction of
XX recombinant vectors that encode PKS enzymes which can be used to make
XX novel polyketides having applications in medicine, agriculture and animal
XX health. The polyketides may have e.g. antibiotic, immunosuppressive or
XX neuroimmunophilin activities
XX
SQ Sequence 640 BP; 92 A; 217 C; 244 G; 87 T; 0 U; 0 Other;
Query Match 4.8%; Score 24; DB 4; Length 640;

Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGGCGCTGCTGGCCACCTACGGCC 466
DB 98 AGGCGCTGCTGGCCACCTACGGCC 121

RESULT 17
AAC55785
ID AAC55785 standard; DNA; 3978 BP.
XX
XX AAC55785;
AC
XX 19-JAN-2001 (first entry)
DT
DE Type I polyketide synthase orf8.
XX
XX Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
XX Streptomyces lavendulae.
XX
XX WO200053737-A2.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US006394.
XX
XX 12-MAR-1999; 99US-00266965.
XX
XX (MINU ) UNIV MINNESOTA.
XX (SHER/) SHERMAN D H.
XX (MAO/) MAO Y.
XX (VARO/) VAROGLU M.
XX (HEMM/) HE M.
XX (SHEL/) SHELDON P C.
XX
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX WPI; 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating
XX the molecular basis of mitosome ring system biosynthesis.
XX
XX Example 1; Page 251-253; 399pp; English.
XX
XX This invention relates to isolated and purified nucleic acid molecules
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
XX natural products that contain a variety of functional groups, including
XX amino benzoquinone and axiridine ring systems. The S. lavendulae
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning
XX 55kb of DNA. The invention includes an expression cassette comprising a
XX mitomycin biosynthetic gene operably linked to a promoter, and host cells
XX transformed with the cassette. The nucleotide, and protein sequences and
XX the transformed host cells of the invention result in antitasthmatic,
XX antiinflammatory, cytostatic, immunomodulatory, and antibiotic
XX activities. The nucleotide sequences are used to elucidate the molecular
XX basis for the biosynthesis of the mitosome ring system, as well as to
XX engineer the biosynthesis of novel natural products, e.g. antibiotics,
XX anti-inflammatory agents, anti-cancer agents, immune-enhancers,
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
XX disease as well as other disease involving respiratory inflammation, or
XX cholesterol-lowering agents or as crop protection agents (e.g. fungicides
XX or insecticides) as well as biopolymers, e.g., in packaging or biomedical
XX applications, or to engineer PHA monomer synthases. Sequences AAC55782-
XX C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin
XX biosynthetic gene cluster DNA sequences and encoded proteins. Sequences
XX AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR
XX primers used in the cloning of the mitomycin biosynthetic genes
```

XX SQ Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 U; 0 Other;

Query Match 4.8%; Score 24; DB 3; Length 3978;  
Best Local Similarity 100.0%; Pred. No. 0.42;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
|||||  
Db 757 CTGGCCCCGACGGCGCTGCAAG 780

## RESULT 18

ADE10204

ID ADE10204 standard; DNA; 3978 BP.

XX AC ADE10204;

XX DT 29-JAN-2004 (first entry)

XX DE S. lavendulae polyketide synthase orfs gene.

XX KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
XX KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.

XX OS Streptomyces lavendulae.

XX PN US2003134398-A1.

XX PD 17-JUL-2003.

XX PF 12-SEP-2001; 2001US-00953348.

XX PR 12-SEP-2001; 2001US-00953348.

XX PA (SHER/) SHERMAN D H.

XX PA (MAOY/) MAO Y.

XX PA (VARO/) VAROGLU M.

XX PA (HEMM/) HE M.

XX PA (SHEL/) SHELDON P.

XX PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;

XX WPI; 2003-863498/80.

XX PT New nucleic acid molecule comprising a sequence having mitomycin  
XX PT biosynthetic gene cluster, useful for enhancing production of  
XX PT antibiotics.

XX PS Example 1; SEQ ID NO 19; 308pp; English.

XX PS The invention relates to an isolated and purified nucleic acid molecule  
XX CC comprising a sequence having mitomycin biosynthetic gene cluster, or its  
XX CC variant or fragment. Also included are an expression cassette comprising  
XX CC the novel nucleic acid molecule operably linked to a promoter functional  
XX CC in a host cell), a recombinant bacterial host cell in which at least a  
XX CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
XX CC cluster is disrupted (resulting in a recombinant host cell that produces  
XX CC altered levels of mitomycin relative to a corresponding nonrecombinant  
XX CC bacterial host cell), introducing exogenous DNA into a refractory  
XX CC Streptomyces strain, identifying a nucleic acid molecule that is related  
XX CC to at least a portion of a nucleic acid molecule comprising a mitomycin  
XX CC gene cluster, preparing a compound or its salt from the recombinant host  
XX CC cell and a product produced by the recombinant host cell. The nucleic  
XX CC acid encodes a MitT, MitS, MitR, MitO, MitP, MitD, MitN, MitM, MitL, and/or  
XX CC MitK, MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or  
XX CC MmCA-MmcyI. The nucleic acid is useful for enhancing production of  
XX CC mitomycin antibiotics, which induce apoptosis and hence are useful as  
XX CC anti-tumour (via tumour hypoxia) agents and are useful in treating  
XX CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The  
XX CC present sequence is a gene encoding a non-mitomycin biosynthetic protein  
XX CC present in the isolated DNA which includes the cluster.

SQ Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 U; 0 Other;

Query Match 4.8%; Score 24; DB 9; Length 3978;  
Best Local Similarity 100.0%; Pred. No. 0.42;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
|||||  
Db 757 CTGGCCCCGACGGCGCTGCAAG 780

## RESULT 19

AAC55840

ID AAC55840 standard; DNA; 12249 BP.

XX AC AAC55840;

XX DT 19-JAN-2001 (first entry)

XX DE Complete Mitomycin ORF 1-9 nucleotide sequence.

XX KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;  
XX KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
XX KW chronic obstructive pulmonary disease; respiratory inflammation;  
XX KW fungicide; pesticide; ds.

XX OS Streptomyces lavendulae.

XX PN WO200053737-A2.

XX PD 14-SEP-2000.

XX PF 10-MAR-2000; 2000WO-US006394.

XX PR 12-MAR-1999; 99US-00266965.

XX PA (MINU ) UNIV MINNESOTA.

XX PA (SHER/) SHERMAN D H.

XX PA (MAOY/) MAO Y.

XX PA (VARO/) VAROGLU M.

XX PA (HEMM/) HE M.

XX PA (SHEL/) SHELDON P C.

XX PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;

XX WPI; 2000-601980/57.

XX PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
XX PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
XX PT the molecular basis of mitosane ring system biosynthesis.

XX PS Claim 26; Fig 22; 399pp; English.

XX PS This invention relates to isolated and purified nucleic acid molecules  
XX CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
XX CC natural products that contain a variety of functional groups, including  
XX CC amino benzoquinone and axiridine ring systems. The S. lavendulae  
XX CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
XX CC 55Kb of DNA. The invention includes an expression cassette comprising a  
XX CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
XX CC transformed with the cassette. The nucleotide, and protein sequences and  
XX CC the transformed host cells of the invention result in antiasthmatic,  
XX CC anti-inflammatory, cytostatic, immunomodulatory, and antibiotic  
XX CC activities. The nucleotide sequences are used to elucidate the molecular  
XX CC basis for the biosynthesis of the mitosane ring system, as well as to  
XX CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
XX CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
XX CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
XX CC disease as well as other disease involving respiratory inflammation, or  
XX CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
XX CC or insecticides) as well as bipolymers, e.g., in packaging or biomedical  
XX CC applications, or to engineer PHA monomer syntheses. Sequences AAC55782-  
XX CC C55681, AAC55815-C55849 and AAB32485-B32542 represent mitomycin

CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
CC primers used in the cloning of the mitomycin biosynthetic genes  
XX  
SQ Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 U; 0 Other;  
Query Match 4.8%; Score 24; DB 3; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
DB 1744 CTGGCCCCGACGGCGCTGCAAG 1767

RESULT 20  
ADE10259  
ID ADE10259 standard; DNA; 12249 BP.  
XX  
AC ADE10259;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE S. lavendulae mitomycin C gene cluster for ORFs 1-9.  
XX  
KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
FN US2003134398-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 12-SEP-2001; 2001US-00953348.  
XX  
PR 12-SEP-2001; 2001US-00953348.  
XX  
PA (SHER/) SHERMAN D H.  
PA (MAO/) MAO Y.  
PA (VARO/) VAROGLU M.  
PA (HEMM/) HE M.  
PA (SHEL/) SHELDON P.  
XX  
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;  
XX  
WPI; 2003-863498/80.  
XX  
DR  
XX  
PT New nucleic acid molecule comprising a sequence having mitomycin  
PT biosynthetic gene cluster, useful for enhancing production of  
PT antibiotics.  
XX  
PS Claim 21; SEQ ID NO 74; 308pp; English.  
XX  
CC The invention relates to an isolated and purified nucleic acid molecule  
CC comprising a sequence having mitomycin biosynthetic gene cluster, or its  
CC variant or fragment. Also included are an expression cassette comprising  
CC the novel nucleic acid molecule (operably linked to a promoter functional  
CC in a host cell), a recombinant bacterial host cell in which at least a  
CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
CC cluster is disrupted (resulting in a recombinant host cell that produces  
CC altered levels of mitomycin relative to a corresponding nonrecombinant  
CC bacterial host cell), introducing exogenous DNA into a refractory  
CC Streptomyces strain, identifying a nucleic acid molecule that is related  
CC to at least a portion of a nucleic acid molecule comprising a mitomycin  
CC gene cluster, preparing a compound or its salt from the recombinant host  
CC cell and a product produced by the recombinant host cell. The nucleic  
CC acid encodes a MitT, MitS, MitR, MitQ, MitP, MitC, MitB, MitL, MitI,  
CC MitK, MitJ, MitI, MitC, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or  
CC MmcA-MmcY. The nucleic acid is useful for enhancing production of  
CC mitomycin antibiotics, which induce apoptosis and hence are useful as  
CC anti-tumour (via tumour hypoxia) agents and are useful in treating  
CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The

CC present sequence is an operon or gene cluster encoding the mitomycin  
CC biosynthetic proteins of the invention.  
XX  
SQ Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 U; 0 Other;  
Query Match 4.8%; Score 24; DB 9; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.36;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
DB 1744 CTGGCCCCGACGGCGCTGCAAG 1767

RESULT 21  
AAC55857  
ID AAC55857 standard; DNA; 18331 BP.  
XX  
AC AAC55857;  
XX  
DT 19-JAN-2001 (first entry)  
XX  
DE Complete nucleotide sequence of the mitomycin gene cluster.  
XX  
KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW fungicide; pesticide; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
FN WO200053737-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 10-MAR-2000; 2000WO-US006394.  
XX  
PR 12-MAR-1999; 99US-00266965.  
XX  
PA (MINU ) UNIV MINNESOTA.  
PA (SHER/) SHERMAN D H.  
PA (MAO/) MAO Y.  
PA (VARO/) VAROGLU M.  
PA (HEMM/) HE M.  
PA (SHEL/) SHELDON P C.  
XX  
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
XX  
WPI; 2000-601980/57.  
XX  
DR  
XX  
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
PT the molecular basis of mitosome ring system biosynthesis.  
XX  
PS Disclosure; Fig 21; 399pp; English.  
XX  
CC This invention relates to isolated and purified nucleic acid molecules  
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
CC natural products that contain a variety of functional groups, including  
CC amino benzoquinone and axiridine ring systems. The S. lavendulae  
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
CC 55kb of DNA. The invention includes an expression cassette comprising a  
CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
CC transformed with the cassette. The nucleotide, and protein sequences and  
CC the transformed host cells of the invention result in antiasthmatic,  
CC antiinflammatory, cytostatic, immunomodulatory, and antibiotic  
CC activities. The nucleotide sequences are used to elucidate the molecular  
CC basis for the biosynthesis of the mitosome ring system, as well as to  
CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
CC disease as well as other disease involving respiratory inflammation, or  
CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides

CC or insecticides) as well as bipolymers, e.g., in packaging or biomedical  
 CC applications, or to engineer PHA monomer synthases. Sequences AAC55782-  
 CC C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin  
 CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
 CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
 CC primers used in the cloning of the mitomycin biosynthetic genes  
 XX  
 SQ Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 U; 0 Other;  
 Query Match 4.8%; Score 24; DB 3; Length 18331;  
 Best Local Similarity 100.0%; Pred. NO. 0.35;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 CTGGCCCCCGACGGCGCTGCAAG 156  
 Db 1744 CTGGCCCCCGACGGCGCTGCAAG 1767  
 RESULT 22  
 ADE10266  
 ID ADE10266 standard; DNA; 18331 BP.  
 XX AC ADE10266;  
 XX DT 29-JAN-2004 (first entry)  
 XX DE  
 XX DE  
 XX S. lavendulae mitomycin gene cluster complete sequence.  
 KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;  
 KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.  
 XX OS Streptomyces lavendulae.  
 XX US2003134398-A1.  
 XX PD 17-JUL-2003.  
 XX 12-SEP-2001; 2001US-00953348.  
 XX 12-SEP-2001; 2001US-00953348.  
 XX (SHER/) SHERMAN D H.  
 XX (MAOY/) MAO Y.  
 XX (VARO/) VAROGLU M.  
 XX (HEMM/) HE M.  
 XX (SHEL/) SHELDON P.  
 XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;  
 WPI; 2003-863498/80.  
 XX New nucleic acid molecule comprising a sequence having mitomycin  
 XX biosynthetic gene cluster, useful for enhancing production of  
 XX antibiotics.  
 XX Claim 29; SEQ ID NO 96; 308pp; English.  
 XX The invention relates to an isolated and purified nucleic acid molecule  
 XX comprising a sequence having mitomycin biosynthetic gene cluster, or its  
 XX variant or fragment. Also included are an expression cassette comprising  
 XX the novel nucleic acid molecule operably linked to a promoter functional  
 XX in a host cell), a recombinant bacterial host cell in which at least a  
 XX portion of a nucleic acid molecule comprising mitomycin biosynthetic gene  
 XX cluster is disrupted (resulting in a recombinant host cell that produces  
 XX altered levels of mitomycin relative to a corresponding nonrecombinant  
 XX bacterial host cell), introducing exogenous DNA into a refractory  
 XX streptomycin strain, identifying a nucleic acid molecule that is related  
 XX to at least a portion of a nucleic acid molecule comprising a mitomycin  
 XX gene cluster, preparing a compound or its salt from the recombinant host  
 XX cell and a product produced by the recombinant host cell. The nucleic  
 XX acid encodes a MitT, MitS, MitR, MitO, MitP, MitN, MitM, MitL,  
 XX MitK, MitU, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitA and/or  
 XX MmCA-Mmcy. The nucleic acid is useful for enhancing production of

CC mitomycin antibiotics, which induce apoptosis and hence are useful as  
 CC anti-tumour (via tumour hypoxia) agents and are useful in treating  
 CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The  
 CC present sequence is an operon or gene cluster encoding the mitomycin  
 CC biosynthetic proteins of the invention.  
 XX  
 SQ Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 U; 0 Other;  
 Query Match 4.8%; Score 24; DB 9; Length 18331;  
 Best Local Similarity 100.0%; Pred. NO. 0.35;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 CTGGCCCCCGACGGCGCTGCAAG 156  
 Db 1744 CTGGCCCCCGACGGCGCTGCAAG 1767  
 RESULT 23  
 ABX04971  
 ID ABX04971 standard; DNA; 103599 BP.  
 XX AC ABX04971;  
 XX DT 16-JAN-2003 (first entry)  
 XX DE  
 XX DE S. cinnamomensis monensin type I polyketide synthase gene cluster.  
 KW Monensin; gene; cluster; polyketide synthase; antibiotic; ds;  
 KW antihelminthic; insecticide; immunosuppressant; antifungal;  
 KW antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;  
 KW mon RI; mon RII; mon T; mon AIX; mon AX.  
 XX OS Streptomyces cinnamomensis.  
 XX WO200168867-A1.  
 XX 20-SEP-2001.  
 XX 30-MAY-2000; 2000WO-GB002072.  
 XX 28-MAY-1999; 99GB-00012563.  
 XX (BIOT-) BIOTICA TECHNOLOGY LTD.  
 XX Leadlay PF, Staunton J, O'Leary M;  
 WPI; 2001-611393/70.  
 XX P-PSDB; ABG99854, ABG99855, ABG99856, ABG99857, ABG99858, ABG99859,  
 XX ABG99860, ABG99861, ABG99862, ABG99863, ABG99864, ABG99865, ABG99866,  
 XX ABG99867, ABG99868, ABG99869, ABG99870, ABG99871, ABG99872, ABG99873,  
 XX ABG99874, ABG99875, ABG99876, ABG99877, ABG99878, ABG99879, ABG99880,  
 XX ABG99881, ABG99882, ABG99883, ABG99884, ABG99885, ABG99886, ABG99887.  
 XX New DNA sequence encoding polyketide synthase, useful for the production  
 XX of polyketides such as antibiotic monensin.  
 XX Claim 1; Page 116-195; 212pp; English.  
 XX The invention relates to a DNA sequence which is a fully defined sequence  
 XX of 103551 base pairs appearing as ABX04971, or its variant, that it is  
 XX not a sequence encoding all or part amino acids 1-920 encoded by mon AI  
 XX as given in the specification. The DNA is the S. cinnamomensis polyketide  
 XX antibiotic monensin biosynthetic gene cluster. Also included are a  
 XX recombinant cloning or expression vector comprising the gene cluster, a  
 XX transformed host cell which has been transformed to contain the gene  
 XX cluster (and is capable of expressing a corresponding polyketide), a  
 XX hybridization probe derived from the gene cluster (for identification and  
 XX isolation of the same or analogous gene cluster, e.g. one which binds  
 XX specifically to a region of the monensin gene cluster selected from mon  
 XX BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX and  
 XX mon AX), the use of the mon RI gene or variant and a monensin promoter to  
 XX control expression of a heterologous gene in Streptomyces cinnamomensis,  
 XX a polypeptide encoded by a portion of the monensin gene cluster



CC (preferably comprising mon BI, mon BII, mon AIX or mon AX or their  
 CC mutants, alleles or variants), an epoxidase enzyme encoded by mon CI, a  
 CC cyclase enzyme encoded by mon CII, producing S. cinnamensis capable of  
 CC enhanced levels of production of monensin comprising engineering it to  
 CC overexpress the mon RI gene. S. cinnamensis containing multiple copies  
 CC of the mon RI gene and/or its variants, expressing a gene heterologous to  
 CC S. cinnamensis comprising transforming S. cinnamensis with DNA  
 CC encoding a heterologous gene and expressing the gene under control of the  
 CC activator gene mon RI or actII/orf4 and 13-propyl erythromycin A. The  
 CC processes and materials (enzyme systems, nucleic acids and vectors) are  
 CC useful for preparing polyketides by recombinant synthesis. The  
 CC polyketides are useful as insecticides, antibiotics, antelmintics,  
 CC antifungals, antibacterials or other pharmaceuticals. In particular the  
 CC gene is useful for the production of monensin, an antibiotic polyether  
 CC polyketide. The present sequence represents the monensin gene cluster  
 XX  
 SQ Sequence 103599 BP; 13980 A; 37023 C; 37799 G; 14795 T; 0 U; 2 Other;  
 Query Match 4.8%; Score 24; DB 4; Length 103599;  
 Best Local Similarity 100.0%; Pred. No. 0.28;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 CTGGCCCCGACGGCGCTGCAAG 156  
 DB 55357 CTGGCCCCGACGGCGCTGCAAG 55380  
 RESULT 24  
 AAA14666  
 ID AAA14666 standard; DNA; 4674 BP.  
 AC AAA14666;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 3..4673  
 FT /tag= a  
 FT /note= "no termination codon given"  
 FT  
 XX  
 PN WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 PF 01-OCT-1999; 99WO-US022886.  
 XX  
 XX 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 PA  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 XX  
 XX WPI: 2000-317716/27.  
 DR P-PSDB; AAY84730.  
 DR  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple

PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 XX Example 2; Page 93-96; 126pp; English.  
 XX  
 CC The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 12 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 SQ Sequence 4674 BP; 704 A; 1873 C; 1464 G; 633 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4674;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCGACGGCGCTGCAACGG 305  
 DB 1254 AACTCGACGGCGCTGCAACGG 1276  
 RESULT 25  
 AAA14665  
 ID AAA14665 standard; DNA; 4725 BP.  
 XX  
 AC AAA14665;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of FK-520 PKS gene cluster module 8.  
 XX  
 KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Streptomyces hygroscopicus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 3..4724  
 FT /tag= a  
 FT /note= "no termination codon given"  
 FT  
 XX  
 PN WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 PF 01-OCT-1999; 99WO-US022886.  
 XX  
 XX 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 PA  
 XX

PI Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 XX WPI; 2000-317716/27.  
 DR P-PSDB; AAY84729.  
 XX  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 PS Example 2; Page 90-93; 126pp; English.  
 XX  
 CC The present sequence encodes module 8 of the FK-520 polyketide synthase  
 CC (PKS) gene cluster of strain MA6548. FK-506 is a potent  
 CC immunosuppressant, and acts through initial formation of an intermediate  
 CC complex with protein immunophilins known as FK-506 binding proteins. The  
 CC nucleic acids are used for producing polyketide compounds. The polyketide  
 CC compounds can be used as immunosuppressants to prevent or treat  
 CC transplant rejection, graft-versus-host disease or uveitis. They can also  
 CC be used for treating e.g. alopecia universalis, autoimmune chronic active  
 CC hepatitis, inflammatory bowel disease, multiple sclerosis, primary  
 CC biliary cirrhosis, or scleroderma. They also have neurotrophic activity  
 CC and can be used to promote neurite outgrowth in NGF-treated PC12 cells  
 CC and in sensory neuronal cultures, and in intact animals, they promote  
 CC regrowth of damaged facial and sciatic nerves, and repair lesioned  
 CC serotonin and dopamine neurons in the brain. They can also be used for  
 CC treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic  
 CC spinal cord and brain injury, or peripheral neuropathies. They can also  
 CC be used in agricultural and veterinary applications  
 XX  
 SQ Sequence 4725 BP; 728 A; 2034 C; 1394 G; 569 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4725;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGGCGCTCGAACGG 305  
 |||||  
 DB 1254 AACTCCGACGGCGCTCGAACGG 1276  
 |||||  
 RESULT 26  
 ID AAA14668 standard; DNA; 4737 BP.  
 AC AAA14668;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 3..4736  
 FT /tag= a  
 FT /note= "no termination codon given"  
 XX  
 XX WO200020601-A2.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 01-OCT-1999; 99WO-US022886.

XX  
 PR 02-OCT-1998; 98US-0102748P.  
 PR 11-MAR-1999; 99US-0123810P.  
 PR 17-JUN-1999; 99US-0139650P.  
 XX  
 XX (KOSA-) KOSAN BIOSCIENCES INC.  
 XX  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 DR WPI; 2000-317716/27.  
 DR P-PSDB; AAY84732.  
 XX  
 PT New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 PS Example 2; Page 99-102; 126pp; English.  
 XX  
 CC The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 12 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 SQ Sequence 4737 BP; 718 A; 1927 C; 1472 G; 620 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 3; Length 4737;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGGCGCTCGAACGG 305  
 |||||  
 DB 1254 AACTCCGACGGCGCTCGAACGG 1276  
 |||||  
 RESULT 27  
 ID AAA14667 standard; DNA; 4767 BP.  
 XX  
 AC AAA14667;  
 XX  
 XX 08-AUG-2000 (first entry)  
 DT  
 DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.  
 XX  
 XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;  
 KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;  
 KW transplant rejection; graft-versus-host disease; alopecia universalis;  
 KW autoimmune chronic active hepatitis; inflammatory bowel disease;  
 KW multiple sclerosis; primary biliary cirrhosis; scleroderma;  
 KW neurite outgrowth; nerve regrowth; Parkinson's disease;  
 KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;  
 KW peripheral neuropathy; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 3..4766



PA (KOSA-) KOSAN BIOSCIENCES INC.  
 XX Reeves C, Chu D, Khosla C, Santi D, Wu K;  
 PI  
 XX  
 XX  
 DR WPI; 2000-317716/27.  
 DR P-PSDB; AAY84733.  
 XX  
 XX New isolated polyketide synthase nucleic acid and polyketide compounds,  
 PT useful for treating e.g. transplant rejection, uveitis, multiple  
 PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or  
 PT peripheral neuropathy.  
 XX  
 XX Example 2; Page 102-105; 126pp; English.  
 PS  
 XX  
 XX The present sequence represents module 8 of the FK-520 polyketide  
 CC synthase (PKS) gene cluster, containing the acyltransferase (AT) domain  
 CC of module 13 of rapamycin. FK-506 is a potent immunosuppressant, and acts  
 CC through initial formation of an intermediate complex with protein  
 CC immunophilins known as FK-506 binding proteins. The nucleic acids are  
 CC used for producing polyketide compounds. The polyketide compounds can be  
 CC used as immunosuppressants to prevent or treat transplant rejection,  
 CC graft-versus-host disease or uveitis. They can also be used for treating  
 CC e.g. alopecia universalis, autoimmune chronic active hepatitis,  
 CC inflammatory bowel disease, multiple sclerosis, primary biliary  
 CC cirrhosis, or scleroderma. They also have neurotrophic activity and can  
 CC be used to promote neurite outgrowth in NGF-treated PC12 cells and in  
 CC sensory neuronal cultures, and in intact animals, they promote regrowth  
 CC of damaged facial and sciatic nerves, and repair lesioned serotonin and  
 CC dopamine neurons in the brain. They can also be used for treating e.g.  
 CC Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord  
 CC and brain injury, or peripheral neuropathies. They can also be used in  
 CC agricultural and veterinary applications  
 XX  
 XX Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 U; 0 Other;  
 SQ Query Match 4.6%; Score 23; DB 3; Length 4818;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 AACTCCGACGCGCGTCAACGG 305  
 DB 1254 AACTCCGACGCGCGTCAACGG 1276  
 RESULT 30  
 ID AAV21186  
 XX AAV21186 standard; DNA; 5676 BP.  
 XX  
 XX  
 AC AAV21186;  
 DT 24-JUL-1998 (first entry)  
 XX  
 XX Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.  
 DE  
 XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
 KW polyketide synthase; actinomycete; ansamycin; ds.  
 XX  
 XX Amycolatopsis mediterranei.  
 OS  
 XX Key Location/Qualifiers  
 FH 3..5676  
 CDS /\*tag= a  
 FT /product= "polyketide synthase"  
 FT /note= "no stop codon given"  
 XX  
 XX WO9807868-A1.  
 XX  
 XX 26-FEB-1998.  
 PD  
 XX 18-AUG-1997; 97WO-EP004495.  
 XX  
 XX 20-AUG-1996; 96EP-00810551.  
 DR  
 XX

PA (NOVS ) NOVARTIS AG.  
 XX  
 XX Schupp T, Toupet C, Engel N;  
 PI  
 XX  
 DR WPI; 1998-169172/15.  
 DR P-PSDB; AAW52844.  
 XX  
 XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
 PT produce rifamycin and rifamycin analogues.  
 PT  
 XX Claim 4; Page 38-43; 205pp; English.  
 PS  
 XX  
 XX The present sequence represents Amycolatopsis mediterranei strain wt3136  
 CC 5.7 kb KpnI fragment DNA, from the present invention. The present  
 CC invention describes a Amycolatopsis mediterranei rifamycin synthesis gene  
 CC cluster DNA fragment comprising a DNA region involved directly or  
 CC indirectly in the gene cluster responsible for rifamycin synthesis,  
 CC including the adjacent DNA regions to the right and left which, by reason  
 CC of their function in connection with rifamycin biosynthesis, qualify as  
 CC constituents of this rifamycin gene cluster, and functional fragments,  
 CC derivatives or constituents of these. The Amycolatopsis mediterranei  
 CC rifamycin synthesis gene cluster DNA fragment can be used for producing  
 CC rifamycin, rifamycin analogues or precursors. It can also be used for  
 CC inactivating or modifying genes involved in ansamycin or rifamycin  
 CC biosynthesis. The DNA can be used for constructing mutant actinomycetes  
 CC strains from which the natural rifamycin or ansamycin biosynthesis gene  
 CC cluster has been partly or completely deleted. The DNA fragment can be  
 CC used for assembling a library of polyketide synthases, which can be used  
 CC for assembling a library of polyketides. A hybridisation probe of the  
 CC invention can be used for identifying DNA fragments involved in the  
 CC biosynthesis of ansamycins  
 XX  
 XX Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 U; 0 Other;  
 SQ Query Match 4.6%; Score 23; DB 2; Length 5676;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 277 GCGGTCAACTCCGACGCGGTC 299  
 DB 4173 GCGGTCAACTCCGACGCGGTC 4195  
 RESULT 31  
 ID AAF88335  
 XX AAF88335 standard; DNA; 7788 BP.  
 XX  
 XX  
 AC AAF88335;  
 DT 28-AUG-2001 (first entry)  
 XX  
 XX S. spinosa DNA fragment encoding ORF18, SEQ ID 41.  
 DE  
 XX Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;  
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
 KW macrocide; insecticidal; polyketide synthase; ds.  
 XX  
 XX Saccharopolyspora spinosa.  
 OS  
 XX DE19957268-A1.  
 PN  
 XX 08-MAR-2001.  
 PD  
 XX 29-NOV-1999; 99DE-01057268.  
 PF  
 XX 27-AUG-1999; 99DE-01040596.  
 PR  
 XX (FARB ) BAYER AG.  
 PA  
 XX Eberz G, Moehrle V, Froede R, Velten R, Salas JA;  
 PI WPI; 2001-267102/28.  
 DR P-PSDB; AAB70965.  
 DR





DR WPI; 2000-565458/52.  
DR P-PSDB; AAB23749, AAB23750.  
XX  
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
PS Claim 2; Page 66-134; 314pp; Japanese.  
XX  
CC The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 3; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||  
RESULT 36  
ID AAH79277 standard; DNA; 30690 BP.  
XX  
AC AAH79277;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.  
XX  
KW Avermectin aglycone synthase; AAS; avermectin derivative;  
KW drug production; veterinary drug; pesticide; ds.  
XX  
OS Streptomyces avermitilis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .11919  
FT /\*tag= a  
FT /product= "AAG65264"  
FT /partial  
FT 11971. .30690  
FT /\*tag= b  
FT /product= "AAG65265"  
XX  
PN WO200162939-A1.  
XX  
PD 30-AUG-2001.  
XX  
PF 23-FEB-2001; 2001WO-JP001381.  
XX  
PR 24-FEB-2000; 2000JP-00047405.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA (KITA ) KITASATO INST.  
XX  
PI Endo H, Yanaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
XX  
DR WPI; 2001-582053/65.  
DR P-PSDB; AAG65264, AAG65265.  
XX  
PT New modified avermectin aglycone synthase derived from Streptomyces

PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
PT drugs and pesticides.  
XX  
PS Example 2; Page 58-123; 257pp; Japanese.  
XX  
CC The present invention relates to the production of modified derivatives  
CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
CC can be used in the production of drugs, veterinary drugs and pesticides.  
CC The present sequence is a fragment of the S. avermitilis genome  
XX  
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||  
RESULT 37  
ID AAA92302 standard; DNA; 31422 BP.  
XX  
AC AAA92302;  
XX  
DT 10-JAN-2001 (first entry)  
XX  
DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.  
XX  
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical; ds.  
XX  
OS Streptomyces avermitilis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .14646  
FT /\*tag= a  
FT /note= "avermectin aglycon synthase protein"  
FT 14824. .31422  
FT /\*tag= b  
FT /note= "avermectin aglycon synthase protein"  
XX  
PN WO200050605-A1.  
XX  
PD 31-AUG-2000.  
XX  
PF 23-FEB-2000; 2000WO-JP001041.  
XX  
PR 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
PI Omura S, Ikeda H;  
XX  
DR WPI; 2000-565458/52.  
DR P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
PS Claim 2; Page 134-203; 314pp; Japanese.  
XX  
CC The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||  
RESULT 37  
ID AAA92302 standard; DNA; 31422 BP.  
XX  
AC AAA92302;  
XX  
DT 10-JAN-2001 (first entry)  
XX  
DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.  
XX  
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical; ds.  
XX  
OS Streptomyces avermitilis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1. .14646  
FT /\*tag= a  
FT /note= "avermectin aglycon synthase protein"  
FT 14824. .31422  
FT /\*tag= b  
FT /note= "avermectin aglycon synthase protein"  
XX  
PN WO200050605-A1.  
XX  
PD 31-AUG-2000.  
XX  
PF 23-FEB-2000; 2000WO-JP001041.  
XX  
PR 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
PI Omura S, Ikeda H;  
XX  
DR WPI; 2000-565458/52.  
DR P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
PS Claim 2; Page 134-203; 314pp; Japanese.  
XX  
CC The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 4; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 418 ACCCTGGCGACCCGATCGAGGC 440  
Db 13015 ACCCTGGCGACCCGATCGAGGC 13037  
|||||





Query Match 4.6%; Score 23; DB 5; Length 33529;  
Best Local Similarity 100.0%; Pred. No. 0.9; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 CCTGGCGACCCGATCGAGCGC 442  
DB 16284 CCTGGCGACCCGATCGAGCGC 16306

RESULT 40  
AAT80414  
ID AAT80414 standard; DNA; 44377 BP.  
XX AC AAT80414;  
XX 27-FEB-1998 (first entry)  
XX Platenolide synthase gene cluster.  
XX Ty lactone synthase gene cluster; tylG gene; multifunctional protein;  
KW platenolide synthase gene cluster; platenolide production; smg gene;  
KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.  
XX OS Streptomyces ambofaciens.  
XX Key Location/Qualifiers  
FT CDS 350..14002  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes protein shown in AAW22606"  
FT CDS 14045..20036  
FT /tag= b  
FT /note= "ORF2 encodes protein shown in AAW22607"  
FT CDS 20110..31284  
FT /tag= c  
FT /transl\_except= (pos:20111..20113, aa:Met)  
FT /note= "ORF3 encodes protein shown in AAW22608"  
FT CDS 31329..36071  
FT /tag= d  
FT /note= "ORF4 encodes protein shown in AAW22609"  
FT CDS 36155..41830  
FT /tag= e  
FT /note= "ORF5 encodes protein shown in AAW22610"  
XX EP791655-A2.  
XX 27-AUG-1997.  
XX 19-FEB-1997; 97EP-00301056.  
XX 22-FEB-1996; 96US-0012078P.  
XX (ELIL ) LILLY & CO ELI.  
XX Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;  
XX WPI; 1997-418046/39.  
XX P-PSDB; AAW22606, AAW22607, AAW22608, AAW22609, AAW22610.  
XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for  
PT production of tylosin-related polyketide compounds.  
XX Example 2; Page 110-134; 220pp; English.  
XX This sequence represents the platenolide synthase gene cluster of the  
CC invention. This sequence is referred to as the smg gene, and was  
CC isolated from Streptomyces ambofaciens. This sequence encodes the multi-  
CC functional proteins which direct the synthesis of the polyketide  
CC platenolide. Platenolide is the basic building block of the macrolide  
CC antibiotic spiramycin. This sequence was used along with the tylG gene  
CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG  
CC gene is the ty lactone synthase gene cluster of the invention. The tylG

CC sequence was isolated from Streptomyces fradiae, and encodes  
CC multifunctional proteins which direct the synthesis of the polyketide  
CC ty lactone. Ty lactone is the basic building block of the antibiotic  
CC tylosin. The hybrid sequence can be used to transform S. ambofaciens  
CC lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1  
CC sequence, so that they can produce polyketides. The DNA sequence can be  
CC modified so as to alter the type of carboxylic acids incorporated, the  
CC number of carboxylic acids incorporated and/or the post-condensation  
CC reactions performed, thereby resulting in novel tylosin-related  
CC polyketides  
XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;  
Query Match 4.6%; Score 23; DB 2; Length 44377;  
Best Local Similarity 100.0%; Pred. No. 0.87;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GAGTTCTCGCGCAGCGGGCT 134  
DB 4157 GAGTTCTCGCGCAGCGGGCT 4179

RESULT 41  
AAT78508  
ID AAT78508 standard; DNA; 44377 BP.  
XX AC AAT78508;  
XX 26-FEB-1998 (first entry)  
XX Platenolide synthase gene cluster.  
XX Platenolide synthase gene cluster; platenolide production; smg gene;  
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.  
XX OS Streptomyces ambofaciens.  
XX Key Location/Qualifiers  
FT CDS 350..14002  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes protein shown in AAW23716"  
FT CDS 14046..20036  
FT /tag= b  
FT /note= "ORF2 encodes protein shown in AAW23717"  
FT CDS 20110..31284  
FT /tag= c  
FT /transl\_except= (pos:20111..20113, aa:Met)  
FT /note= "ORF3 encodes protein shown in AAW23718"  
FT CDS 31329..36071  
FT /tag= d  
FT /note= "ORF4 encodes protein shown in AAW23719"  
FT CDS 36155..41830  
FT /tag= e  
FT /note= "ORF5 encodes protein shown in AAW23720"  
XX EP791656-A2.  
XX 27-AUG-1997.  
XX 19-FEB-1997; 97EP-00301066.  
XX 22-FEB-1996; 96US-0012050P.  
XX (ELIL ) LILLY & CO ELI.  
XX Burgett SG, Kuhstoss SA, Rao RN, Richardson MA, Rosteck PR;  
XX WPI; 1997-418047/39.  
XX P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.  
XX DNA encoding Streptomyces ambofaciens platenolide synthase domain - for  
PT production of spiramycin-related polyketide antibiotics.

XX Claim 9; Page 8-33; 81pp; English.

XX This sequence represents the platenolide synthase gene cluster of the

XX invention. This sequence is referred to as the *srmd* gene, and was

XX isolated from *Streptomyces ambofaciens*. This sequence encodes the multi-

XX functional proteins which direct the synthesis of the polyketide

XX platenolide. Platenolide is the basic building block of the macrolide

XX antibiotic spiramycin. The DNA can be used to produce compounds

XX exhibiting antibiotic activity based on the platenolide structure,

XX including specifically the macrolide antibiotic spiramycin and spiramycin

XX analogues and derivatives. Modifications of the platenolide synthase DNA

XX sequence can be made so as to change the number and type of carboxylic

XX acids incorporated into the growing polyketide chain and to change the

XX kind of post-condensation processing that is conducted

XX

XX Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 2; Length 44377;

Best Local Similarity 100.0%; Pred.No. 0.87;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GAGTTCGCGCGCAGCGGGCT 134

DB 4157 GAGTTCGCGCGCAGCGGGCT 4179

RESULT 42

AAF88312

ID AAF88312 standard; DNA; 50000 BP.

XX AAF88312;

XX

DT 28-AUG-2001 (first entry)

XX

DE S. spinosa DNA fragment SEQ ID 1.

XX

KW Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;

KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;

KW macrolide; insecticidal; ds.

XX

OS Saccharopolyspora spinosa.

XX

PN DE19957268-A1.

XX

PD 08-MAR-2001.

XX

PF 29-NOV-1999; 99DE-01057268.

XX

PR 27-AUG-1999; 99DE-01040596.

XX

PA (FARB ) BAYER AG.

XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

XX

DR WPI; 2001-267102/28.

XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

PT recombinant production of insecticidal spinosyns and their derivatives.

XX

PS Claim 7; Page 14-31; 354pp; German.

XX

CC This invention describes a novel method nucleic acid (I) and its encoded

CC polypeptide (II) containing at least one region that encodes an enzymatic

CC activity involved in biosynthesis of spinosyns. (I) are used (i) to

CC identify, inactivate or modulate genes involved in the biosynthesis of

CC (II); (ii) to generate a library of polyketide synthases; (iii) for

CC adding forosamine or trimethylrharnnose to a spinosyn or polyketide

CC aglycone; and (iv) for recombinant production of the corresponding

CC enzymes, which are used for production of (II), their precursors or

CC derivatives, including production of transgenic plants that express (II)

CC and thus have increased resistance to insects. (I) are also useful as

CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

CC macrolides with insecticidal, but not antibacterial, activity, and can

CC also be used to raise specific antibodies, useful for identifying

CC expression clones in a gene bank. Cells transformed with (I) may produce

CC (II) at significantly increased levels or produce new derivatives of

CC (II). This sequence represents a genomic DNA fragment of the *S. spinosa*

CC genome which contains the coding regions for proteins involved in

CC forosamine, trimethylrharnnose and polyketide synthase biosynthesis

XX

XX Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 4; Length 50000;

Best Local Similarity 100.0%; Pred.No. 0.85;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGCACCGGCTGCTCGAGGG 197

DB 49400 GACGGCACCGGCTGCTCGAGGG 49422

RESULT 43

AAF88316

ID AAF88316 standard; DNA; 50000 BP.

XX AAF88316;

XX

DT 28-AUG-2001 (first entry)

XX

DE S. spinosa DNA fragment SEQ ID 5.

XX

KW Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;

KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;

KW macrolide; insecticidal; ds.

XX

OS Saccharopolyspora spinosa.

XX

PN DE19957268-A1.

XX

PD 08-MAR-2001.

XX

PF 29-NOV-1999; 99DE-01057268.

XX

PR 27-AUG-1999; 99DE-01040596.

XX

PA (FARB ) BAYER AG.

XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

XX

DR WPI; 2001-267102/28.

XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

PT recombinant production of insecticidal spinosyns and their derivatives.

XX

PS Claim 7; Page 74-91; 354pp; German.

XX

CC This invention describes a novel method nucleic acid (I) and its encoded

CC polypeptide (II) containing at least one region that encodes an enzymatic

CC activity involved in biosynthesis of spinosyns. (I) are used (i) to

CC identify, inactivate or modulate genes involved in the biosynthesis of

CC (II); (ii) to generate a library of polyketide synthases; (iii) for

CC adding forosamine or trimethylrharnnose to a spinosyn or polyketide

CC aglycone; and (iv) for recombinant production of the corresponding

CC enzymes, which are used for production of (II), their precursors or

CC derivatives, including production of transgenic plants that express (II)

CC and thus have increased resistance to insects. (I) are also useful as

CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

```
SQ Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 U; 0 Other;
Query Match 4.6%; Score 23; DB 4; Length 50000;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGCACCGCTGCTCCGAGGG 197
|||||
Db 3776 GACGGCACCGCTGCTCCGAGGG 3798

RESULT 44
AADI17184
ID AADI17184 standard; DNA; 65140 BP.
XX AC
XX AADI17184;
XX 29-NOV-2001 (first entry)
XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nys1; ds.
XX Streptomyces noursei.
XX Key Location/Qualifiers
FH CDS complement(1..1035)
FT /*tag= a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT CDS complement(1056..2576)
FT /*tag= b
FT /product= "NysD1 protein"
FT 2806..6906
FT /*tag= c
FT /product= "NysA protein"
FT 6952..16530
FT /*tag= d
FT /product= "NysB protein"
FT 16550..49840
FT /*tag= e
FT /product= "NysC protein"
FT 50260..51015
FT /*tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /*tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /*tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT CDS 57180..59963
FT /*tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /*tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT CDS 61736..62497
FT /*tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT CDS complement(62551..63615)
FT /*tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT CDS 63765..64961
FT /*tag= m
FT /product= "ORF1 protein"
XX WO200159126-A2.
```

```
XX 16-AUG-2001.
XX PD
XX 08-FEB-2001; 2001WO-GB0000509.
XX PF
XX 08-FEB-2000; 2000GB-00002840.
XX PR
XX 10-APR-2000; 2000GB-00008786.
XX PR
XX 14-APR-2000; 2000GB-00009387.
XX XX
XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTERF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (EJAE/) EJAERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX (VALL/) VALLA S.
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX as antibiotics and antifungals.
XX Claim 2; Page 116-151; 266pp; English.
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin. The
XX nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster
XX SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 U; 0 Other;
Query Match 4.6%; Score 23; DB 4; Length 65140;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 AACGCCCGTCCAGCAGCGGT 341
|||||
Db 12523 AACGCCCGTCCAGCAGCGGT 12545

RESULT 45
ADC26995
ID ADC26995 standard; DNA; 67251 BP.
XX AC
XX ADC26995;
XX 18-DEC-2003 (first entry)
XX XX
XX Sorangium cellulosum tmbA gene cluster.
XX recombinant DNA vector; tmbA gene cluster; tmbA; tmbB; tmbC;
XX polyketide synthase enzyme; PKS enzyme; hybrid PKS enzyme; tombamycin;
XX ds.
XX Polyangium cellulosum.
XX OS
XX US2003054547-A1.
XX PN
XX 20-MAR-2003.
XX PD
XX 28-AUG-2001; 2001US-00942025.
XX PF
XX 22-JAN-1998; 98US-00010809.
XX PR
```

```

PR 31-AUG-1998; 98US-00144085.
PR 15-FEB-2001; 2001US-0271245P.
XX
XX PA (JULI/) JULIEN B.
XX
XX PI Julien B;
XX
XX WPI; 2003-567100/53.
XX
XX New recombinant DNA vector for preparing hybrid polyketide synthase
XX PT enzymes comprises all or a portion of one or more genes in the tmBA gene
XX PT cluster of Sorangium cellulosum.
XX
XX PS Disclosure; SEQ ID NO 1; 73pp; English.
XX
XX CC The invention describes a recombinant DNA vector that comprises all or a
XX CC portion of one or more genes in the tmBA gene cluster. Also disclosed is
XX CC the host cell that comprises the tmBA genes of Sorangium cellulosum or
XX CC fragments of those genes. The DNA vector is useful in preparing hybrid
XX CC polyketide synthase (PKS) enzymes and the polyketides produced by such
XX CC hybrid enzymes. The gene products of the tmBA cluster can be used to
XX CC synthesize the polyketide tombamycin. This sequence represents the
XX CC Sorangium cellulosum tmBA gene cluster.
XX
XX SQ Sequence 67251 BP; 9090 A; 19855 C; 27880 G; 10420 T; 0 U; 6 Other;
XX
XX Query Match 4.6%; Score 23; DB 9; Length 67251;
XX Best Local Similarity 100.0%; Pred. No. 0.82;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 420 CCTGGGCGACCGATCGAGGCGG 442
DB 32521 CCTGGGCGACCGATCGAGGCGG 32543

```

Search completed: June 16, 2004, 17:43:54  
Job time : 250 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 Seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501  
Sequence: 1 tcgtcttcgtgctgcgct.....cgctgtgctgcggtcggtg 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 759

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS-COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	36	7.2	43280	2	US-08-804-227C-1
2	32	6.4	20235	1	US-07-642-734C-3
3	32	6.4	20235	3	US-08-439-009A-3
4	29	5.8	77536	4	US-09-410-551B-1
5	27	5.4	15872	3	US-09-105-537-1
6	27	5.4	15872	4	US-09-091-609-1
7	27	5.4	15872	4	US-09-091-609-3
8	25	5.0	50937	3	US-09-428-517-1
9	24	4.8	640	3	US-09-154-083-3
10	24	4.8	3978	4	US-09-266-965-19
11	24	4.8	12249	4	US-09-266-965-74
12	24	4.8	18331	4	US-09-266-965-96
13	23	4.6	4674	4	US-09-410-551B-26
14	23	4.6	4725	4	US-09-410-551B-24
15	23	4.6	4737	4	US-09-410-551B-30
16	23	4.6	4767	4	US-09-410-551B-28
17	23	4.6	4818	4	US-09-410-551B-32
18	23	4.6	11219	1	US-07-642-734C-1
19	23	4.6	11219	3	US-08-439-009A-1
20	23	4.6	13987	2	US-08-804-227C-13
21	23	4.6	33529	3	US-09-144-085-3
22	23	4.6	44377	2	US-08-804-227C-7
23	23	4.6	44377	2	US-08-804-198-1
24	23	4.6	80161	3	US-09-036-987A-1
25	23	4.6	80161	3	US-09-370-700-1
26	23	4.6	80161	4	US-09-603-207-1
27	22	4.4	591	3	US-09-010-809-6

c	28	21	4.2	403	3	US-09-060-756-391
c	29	21	4.2	403	4	US-09-670-314-391
	30	21	4.2	13842	3	US-09-105-537-30
	31	21	4.2	36778	3	US-09-105-537-5
	32	21	4.2	38506	3	US-09-320-878-19
	33	21	4.2	38506	4	US-09-141-908-1
	34	21	4.2	38506	4	US-09-657-440-19
	35	20	4.0	258	4	US-09-252-991A-2488
c	36	20	4.0	699	1	US-08-485-721-1
c	37	20	4.0	699	2	US-08-392-935-1
c	38	20	4.0	699	3	US-08-897-236-1
c	39	20	4.0	699	3	US-09-167-874-1
c	40	20	4.0	699	4	US-09-500-253B-1
c	41	20	4.0	699	5	PCT-US93-08326-1
c	42	20	4.0	1180	3	US-08-897-236-10
c	43	20	4.0	1180	3	US-09-167-874-10
c	44	20	4.0	1180	4	US-09-500-253B-10
c	45	20	4.0	1332	4	US-09-252-991A-4685

ALIGNMENTS

RESULT 1  
US-08-804-227C-1  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuntz, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43280 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 816..14234  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14351..19945  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20010..31199

Sequence 391, Appl  
Sequence 391, Appl  
Sequence 30, Appl  
Sequence 5, Appl  
Sequence 19, Appl  
Sequence 1, Appl  
Sequence 19, Appl  
Sequence 2488, Ap  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 4685, Ap

```
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1

Query March 7.2%; Score 36; DB 2; Length 43280;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGCCCGCCGACGGCGCTGCAAGCGCTTCGCCGAC 168
Db 15074 CTGCCCGCCGACGGCGCTGCAAGCGCTTCGCCGAC 15109

RESULT 2
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dancickers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
; OTHER INFORMATION: 6-deoxyerythronolide B formatio"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximate span of
```

```

; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
;
; US-07-642-734C-3
;
; Query Match 6.4%; Score 32; DB 1; Length 20235;
; Best Local Similarity 100.0%; Pred. No. 9,1e-06;
; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 382 GACGTCGACGCGGTGGAGCGCACGCCGCGG 413
; |||||
; DB 1006 GACGTCGACGCGGTGGAGCGCACGCCGCGG 1037
; |||||
;
; RESULT 3
; US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, I
; APPLICANT: Meallpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:

```

```
NAME/KEY: misc feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formation"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc feature
```

```
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3
Query Match 6.4%; Score 32; DB 3; Length 20235;
Best Local Similarity 100.0%; Pred. No. 9.1e-06; Indels 0; Gaps 0;
Matches 32; Conservative 0; Mismatches 0;
QY 382 GAGCTGACGCGGTGGAGCGCACGCGCACGCGG 413
| | | | | | | | | | | | | | | | | | | |
Db 1006 GAGCTGACGCGGTGGAGCGCACGCGCACGCGG 1037
| | | | | | | | | | | | | | | | | | | |
RESULT 4
US-09-410-551B-1/c
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 77536
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275) ... (71465)
US-09-410-551B-1
Query Match 5.8%; Score 29; DB 4; Length 77536;
Best Local Similarity 100.0%; Pred. No. 0.00018; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 0;
QY 277 GGGGTCAACTCGACGCGCGGTGCGAACGG 305
| | | | | | | | | | | | | | | | | | | |
Db 37516 GGGGTCAACTCGACGCGCGGTGCGAACGG 37488
| | | | | | | | | | | | | | | | | | | |
RESULT 5
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
```



```

; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match      5.4%; Score 27; DB 3; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 3747

RESULT 6
US-09-091-609-1
; Sequence 1, Application US/09091609
; Patent No. 6600029
; GENERAL INFORMATION:
; APPLICANT: SHERMAN, DAVID H.
; APPLICANT: WILLIAMS, MARK D.
; APPLICANT: XUE, YONGQUAN
; TITLE OF INVENTION: METABOLIC ENGINEERING OF
; POLYHYDROXYALKANOATE MONOMER SYNTHASES
; FILE REFERENCE: 600.297US2
; CURRENT APPLICATION NUMBER: US/09/091,609
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: PCT/US96/20119
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: 60/008,847
; EARLIER FILING DATE: 1995-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (20)...(13909)
US-09-091-609-1

Query Match      5.4%; Score 27; DB 4; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 3747

RESULT 7
US-09-091-609-3
; Sequence 3, Application US/09091609
; Patent No. 6600029
; GENERAL INFORMATION:
; APPLICANT: SHERMAN, DAVID H.
; APPLICANT: WILLIAMS, MARK D.
; APPLICANT: XUE, YONGQUAN
; TITLE OF INVENTION: METABOLIC ENGINEERING OF
; POLYHYDROXYALKANOATE MONOMER SYNTHASES
; FILE REFERENCE: 600.297US2
; CURRENT APPLICATION NUMBER: US/09/091,609
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: PCT/US96/20119
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: 60/008,847
; EARLIER FILING DATE: 1995-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae

; TYPE: DNA
; NAME/KEY: CDS
; LOCATION: (14148)...(15824)
US-09-091-609-3

Query Match      5.4%; Score 27; DB 4; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGCGCAGCGCGG 3747

RESULT 8
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match      5.0%; Score 25; DB 3; Length 50937;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GAGGCGCGCTGCAAGCGGTTCGCGG 166
Db 9582 GAGGCGCGCTGCAAGCGGTTCGCGG 9606

RESULT 9
US-09-154-083-3
; Sequence 3, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-3

Query Match      4.8%; Score 24; DB 3; Length 640;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Query Match 4.8%; Score 24; DB 4; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.042;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 RESULT 13
2 US-09-410-551B-26
3 US-09-410-551B-26
4 Sequence 26, Application US/09410551B
5 Patent No. 6503737
6 GENERAL INFORMATION:
7 APPLICANT: KOSAN BIOSCIENCES, Inc.
8 APPLICANT: REEVES, CHRISTOPHER
9 APPLICANT: CHU, DANIEL
10 APPLICANT: KHOSLA, CHAITAN
11 APPLICANT: SANTI, DANIEL
12 APPLICANT: WU, KAI
13 TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
14 FILE REFERENCE: 30062-20026.00
15 CURRENT APPLICATION NUMBER: US/09/410,551B
16 CURRENT FILING DATE: 1999-10-01
17 PRIOR APPLICATION NUMBER: US 60/139,650
18 PRIOR FILING DATE: 1999-06-17
19 PRIOR APPLICATION NUMBER: US 60/123,810
20 PRIOR FILING DATE: 1999-03-11
21 PRIOR APPLICATION NUMBER: US 60/102,748
22 PRIOR FILING DATE: 1998-10-02
23 NUMBER OF SEQ ID NOS: 72
24 SOFTWARE: FastSeq for Windows Version 4.0
25 SEQ ID NO 26
26 LENGTH: 4674
27 TYPE: DNA
28 ORGANISM: Artificial Sequence
29 FEATURE:
30 OTHER INFORMATION: Description of Artificial Sequence: DNA encoding

```

```

; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4673)
US-09-410-551B-26

Query Match
Best Local Similarity 100.0%; DB 4; Length 4674;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGTCTGAACGG 305
Db 1254 AACTCCGACGGCGTCTGAACGG 1276

RESULT 14
US-09-410-551B-24
; Sequence 24, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4724)
US-09-410-551B-24

Query Match
Best Local Similarity 100.0%; DB 4; Length 4725;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGTCTGAACGG 305
Db 1254 AACTCCGACGGCGTCTGAACGG 1276

RESULT 15
US-09-410-551B-30
; Sequence 30, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4724)
US-09-410-551B-24

Query Match
Best Local Similarity 100.0%; DB 4; Length 4725;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGTCTGAACGG 305
Db 1254 AACTCCGACGGCGTCTGAACGG 1276

RESULT 16
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
US-09-410-551B-28

Query Match
Best Local Similarity 100.0%; DB 4; Length 4767;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGTCTGAACGG 305

```

Db 1254 AACTCCGACGGCGGTGACGG 1276

RESULT 17

US-09-410-551B-32  
; Sequence 32, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE OF INVENTION: CONSTRUCTS THEREFOR  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 4818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(4817)  
US-09-410-551B-32

Query Match 4.6%; Score 23; DB 4; Length 4818;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 283 AACTCCGACGGCGGTGACGG 305  
Db 1254 AACTCCGACGGCGGTGACGG 1276

RESULT 18

US-07-642-734C-1  
; Sequence 1, Application US/07642734C  
; Patent No. 5824513  
; GENERAL INFORMATION:  
; APPLICANT: Katz, L  
; APPLICANT: Donadio, S  
; APPLICANT: Mcalpine, J B  
; TITLE OF INVENTION: Recombinant DNA Method for Producing  
; TITLE OF INVENTION: Erythromycin Analogs  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Edward H. Gorman  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; STREET: Park Rd  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckerts, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 2338  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 744..6659  
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF  
OTHER INFORMATION: MODULE 1"  
OTHER INFORMATION: /label= FUNCTION  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 744..11219  
OTHER INFORMATION: /function= "eryA"  
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 744..1868  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1998..2198  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2250..3626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3831..4811  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase 2 domain of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5574..6125  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6369..6626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 2 of module 1"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6678..11219  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 2"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6678..8066

Thu Jun 17 09:02:24 2004

OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACPSynthase of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9906..10454  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-07-642-734C-1

Query Match 4.6%; Score 23; DB 1; Length 11219;  
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0;

QY 275 CGGCGTCAACTCCGACGCGCG 297  
Db 3049 CGGCGTCAACTCCGACGCGCG 3071

RESULT 19

US-08-439-009A-1  
Sequence 1, Application US/08439009A  
Patent No. 6004787  
GENERAL INFORMATION:  
APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Steven F. Weinstein  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Park Rd  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea

STRAIN: NRRL 2338  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 744..6659  
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF  
OTHER INFORMATION: MODULE 1"  
OTHER INFORMATION: /label= FUNCTION  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 744..11219  
OTHER INFORMATION: /function= "gene= "eryA"  
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 744..1868  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1998..2198  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 2250..3626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3831..4811  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase 2 domain of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 5574..6125  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6369..6626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 2 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6678..11219  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6678..8066  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACPSynthase of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9906..10454  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-08-439-009A-1

Query Match 4.6%; Score 23; DB 3; Length 11219;  
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0; Indels 0;

Thu Jun 17 09:02:24 2004

```

; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match      4.6%; Score 23; DB 3; Length 33529;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      420 CCTGGCGACCCGATCGAGCGC 442
Db      16284 CCTGGCGACCCGATCGAGCGC 16306

RESULT 22
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII (DOS) Text only
; CURRENT APPLICATION DATA:
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
US-08-804-227C-13

Query Match      4.6%; Score 23; DB 2; Length 13987;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      112 GAGTTCTCGCGCAGCGCGGCT 134
Db      4142 GAGTTCTCGCGCAGCGCGGCT 4164

RESULT 21
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20

```

FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7
Query Match
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Length 44377;
112 GAGTTCTCGCGGCGAGCGGGCT 134
4157 GAGTTCTCGCGGCGAGCGGGCT 4179
RESULT 23
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuntz, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELEPHONE: 317-276-3885
TELEFAX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1
Query Match
4.6%; Score 23; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
112 GAGTTCTCGCGGCGAGCGGGCT 134
4157 GAGTTCTCGCGGCGAGCGGGCT 4179
RESULT 24
US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1
Query Match
4.6%; Score 23; DB 3; Length 80161;
Best Local Similarity 100.0%; Pred. No. 0.098;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
175 GACGGCACCGGCTGGTCGAGGG 197
24771 GACGGCACCGGCTGGTCGAGGG 24793
RESULT 25
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.

; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/09/370,700  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/36987  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-1

Query Match 4.6%; Score 23; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 GACGGCACCGGCTGTCGAGGG 197  
Db 24771 GACGGCACCGGCTGTCGAGGG 24793

RESULT 26  
US-09-603-207-1  
; Sequence 1, Application US/09603207B  
; Patent No. 6521406  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/09/603,207B  
; CURRENT FILING DATE: 2000-06-23  
; EARLIER APPLICATION NUMBER: 09/370,700  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-603-207-1

Query Match 4.6%; Score 23; DB 4; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 175 GACGGCACCGGCTGTCGAGGG 197  
Db 24771 GACGGCACCGGCTGTCGAGGG 24793

RESULT 27  
US-09-010-809-6  
; Sequence 6, Application US/09010809B  
; Patent No. 6090601  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C  
; TITLE OF INVENTION: Epothilone Polyketide Synthases and Encoding DNA  
; TITLE OF INVENTION: Therefor  
; FILE REFERENCE: 30062-2020.00  
; CURRENT APPLICATION NUMBER: US/09/010,809B

; CURRENT FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 591  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-010-809-6

Query Match 4.4%; Score 22; DB 3; Length 591;  
Best Local Similarity 100.0%; Pred. No. 0.46;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 386 TCGACGGCGGTGGAGCGCACGG 407  
Db 233 TCGACGGCGGTGGAGCGCACGG 254

RESULT 28  
US-09-060-756-391/c  
; Sequence 391, Application US/09060756  
; Patent No. 6183957  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/060,756  
; CURRENT FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 391  
; LENGTH: 403  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (various positions within the sequence)  
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"  
US-09-060-756-391

Query Match 4.2%; Score 21; DB 3; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 393 GGTGGAGCGCGCACGGCACCGG 413  
Db 383 GGTGGAGCGCGCACGGCACCGG 363

RESULT 29  
US-09-670-314-391/c  
; Sequence 391, Application US/09670314  
; Patent No. 6492506  
; GENERAL INFORMATION:  
; APPLICANT: Cole, Stewart  
; APPLICANT: Buchrieser-Brosch, Roland  
; APPLICANT: Gordon, Stephen  
; APPLICANT: Billault, Alain  
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
; FILE REFERENCE: 3495-0169  
; CURRENT APPLICATION NUMBER: US/09/670,314  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 09/060,756  
; PRIOR FILING DATE: 1998-04-16  
; NUMBER OF SEQ ID NOS: 743  
; SOFTWARE: PatentIn Ver. 2.0



```
; SEQ ID NO 391
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-670-314-391

Query Match
Best Local Similarity 4.2%; Score 21; DB 4; Length 403;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 GGTGGAGGCGCACGGCACCGG 413
Db 383 GGTGGAGGCGCACGGCACCGG 363

RESULT 30
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 13842;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 1202 TGGGCGACCCGATCGAGGCGG 1222

RESULT 31
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 36778;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 2943 TGGGCGACCCGATCGAGGCGG 2963

RESULT 32
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match
Best Local Similarity 4.2%; Score 21; DB 3; Length 38506;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 TGGGCGACCCGATCGAGGCGG 442
Db 1085 TGGGCGACCCGATCGAGGCGG 1105

RESULT 33
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
```

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-141-908-1

Query Match 4.2%; Score 21; DB 4; Length 38506;  
Best Local Similarity 100.0%; Pred.No. 0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 TGGCGACCCGATCGAGCCG 442

Db 1085 TGGCGACCCGATCGAGCCG 1105

RESULT 34

US-09-657-440-19

; Sequence 19, Application US/09657440

; Patent No. 6509455

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/657,440

; CURRENT FILING DATE: 2000-09-07

; PRIOR APPLICATION NUMBER: 09/320,878

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: CIP OF 09/141,908

; PRIOR FILING DATE: 1998-08-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 19

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-657-440-19

Query Match 4.2%; Score 21; DB 4; Length 38506;  
Best Local Similarity 100.0%; Pred.No. 0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 TGGCGACCCGATCGAGCCG 442

Db 1085 TGGCGACCCGATCGAGCCG 1105

RESULT 35

US-09-252-991A-2488

; Sequence 2488, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2488

; LENGTH: 258

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2488

Query Match

Best Local Similarity 4.0%; Score 20; DB 4; Length 258;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 446 CGTGTGGCCACTACGGC 465

Db 131 CGTGTGGCCACTACGGC 150

RESULT 36

US-08-485-721-1/c

; Sequence 1, Application US/08485721

; Patent No. 5821124

; GENERAL INFORMATION:

; APPLICANT: Regeneron Pharmaceuticals, Inc. and

; APPLICANT: Regents of the University of California

; TITLE OF INVENTION: Dorsal Tissue Affecting Factor and

; TITLE OF INVENTION: Compositions

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Regeneron Pharmaceuticals, Inc.

; STREET: 777 Old Saw Mill River Road

; CITY: Tarrytown

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10591

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/485,721

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/392,935

; FILING DATE: 02-SEP-1993

; APPLICATION NUMBER: PCT/US93/08326

; FILING DATE: 02-SEP-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Kempler Ph.D., Gail M.

; REGISTRATION NUMBER: 32,143

; REFERENCE/DOCKET NUMBER: Reg 132

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 914-347-7000

; TELEFAX: 914-347-2113

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 699 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..699

US-08-485-721-1

Query Match 4.0%; Score 20; DB 1; Length 699;  
Best Local Similarity 100.0%; Pred.No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 TGGCCCCCGACGCCGCTGC 153

Db 346 TGGCCCCCGACGCCGCTGC 327

RESULT 37

US-08-392-935-1/c

; Sequence 1, Application US/08392935

; Patent No. 5843775

; GENERAL INFORMATION:

APPLICANT: Regeneron Pharmaceuticals, Inc. and  
APPLICANT: Regents of the University of California  
TITLE OF INVENTION: Dorsal Tissue Affecting Factor and  
TITLE OF INVENTION: Compositions  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,935  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08326  
FILING DATE: 02-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler Ph.D., Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: Reg 132  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-347-7000  
TELEFAX: 914-347-2113  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 699 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..699  
US-08-392-935-1

Query Match 4.0%; Score 20; DB 2; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
Db 346 TGGCCCCCGACGGCGCTGC 327

RESULT 38  
US-08-897-236-1/c  
Sequence 1, Application US/08897236A  
Patent No. 6075007  
GENERAL INFORMATION:  
APPLICANT: Regeneron Pharmaceuticals, Inc.  
TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
FILE REFERENCE: REG 133  
CURRENT APPLICATION NUMBER: US/08/897,236A  
CURRENT FILING DATE: 1997-07-17  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-08-897-236-1

Query Match 4.0%; Score 20; DB 3; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
Db 346 TGGCCCCCGACGGCGCTGC 327

RESULT 39  
US-09-167-874-1/c  
Sequence 1, Application US/09167874  
Patent No. 6277593  
GENERAL INFORMATION:  
APPLICANT: Valenzuela et al.  
TITLE OF INVENTION: DORSAL TISSUE AFFECTING FACTOR AND COMPOSITIONS  
FILE REFERENCE: REG132-B  
CURRENT APPLICATION NUMBER: US/09/167,874  
CURRENT FILING DATE: 1998-10-07  
EARLIER APPLICATION NUMBER: 08/485,721  
EARLIER FILING DATE: 1995-07-06  
EARLIER APPLICATION NUMBER: 08/392,935  
EARLIER FILING DATE: 1995-09-22  
EARLIER APPLICATION NUMBER: PCT/US93/08326  
EARLIER FILING DATE: 1993-09-02  
EARLIER APPLICATION NUMBER: 07/957,401  
EARLIER FILING DATE: 1992-10-06  
EARLIER APPLICATION NUMBER: 07/950,410  
EARLIER FILING DATE: 1992-09-23  
EARLIER APPLICATION NUMBER: 07/939,954  
EARLIER FILING DATE: 1992-09-03  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-09-167-874-1

Query Match 4.0%; Score 20; DB 3; Length 699;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
Db 346 TGGCCCCCGACGGCGCTGC 327

RESULT 40  
US-09-500-253B-1/c  
Sequence 1, Application US/09500253B  
Patent No. 6500640  
GENERAL INFORMATION:  
APPLICANT: Regeneron Pharmaceuticals, Inc.  
TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
FILE REFERENCE: REG 133-Z  
CURRENT APPLICATION NUMBER: US/09/500,253B  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 699  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(696)  
US-09-500-253B-1

Query Match 4.0%; Score 20; DB 4; Length 699;

```

; Patent No. 6075007
;
; GENERAL INFORMATION:
;
; APPLICANT: Regeneron Pharmaceuticals, Inc.
; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition
; FILE REFERENCE: REG 133
; CURRENT APPLICATION NUMBER: US/08/897,236A
; CURRENT FILING DATE: 1997-07-17
;
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10

```

Qy	134	TGGCCCCCGACGGCCGCTGC	153
Db	766	TGGCCCCCGACGGCCGCTGC	747

RESULT 44  
US-09-500-253B-10/c  
; Sequence 10, Application US/09500253B  
; Patent No. 6500640  
; GENERAL INFORMATION:  
; APPLICANT: Regeneron Pharmaceuticals, Inc.  
; TITLE OF INVENTION: Modified Dorsal Tissue Affecting Factor and Composition  
; FILE REFERENCE: REG 133-Z  
; CURRENT APPLICATION NUMBER: US/09/500,253B  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1180  
; TYPE: DNA  
; ORGANISM: Mouse  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (421)..(1161)  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (16)..(16)  
; OTHER INFORMATION: n = a,c,g, or t  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (235)..(235)  
; OTHER INFORMATION: n = a,c,g, or t  
US-09-500-253B-10

Query Match 4.0%; Score 20; DB 4; Length 1180;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGGCCCCCGACGGCGCTGC 153  
|||||  
Db 766 TGGCCCCCGACGGCGCTGC 747

RESULT 45  
US-09-252-991A-4685/c  
; Sequence 4685, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4685  
; LENGTH: 1332  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4685

Query Match 4.0%; Score 20; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 TGCTGGCCACCTACGGCCAG 468  
|||||  
Db 1283 TGCTGGCCACCTACGGCCAG 1264

Search completed: June 16, 2004, 19:32:38  
Job time : 60.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)  
9488.873 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tgcgttcgctgctgcgct.....cgctggtcggtcggtcggtg 501

Scoring table:

Gapop\_60.0 , Gapext 60.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 3230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	6.4	84428	17	US-10-229-148B-1 Sequence 1, Appli
2	29	5.8	18435	15	US-10-156-761-412 Sequence 412, Appl
3	29	5.8	77536	10	US-09-940-316B-1 Sequence 1, Appli
4	29	5.8	100000	15	US-10-156-761-15103 Sequence 15103, A
5	29	5.8	9025608	15	US-10-156-761-1 Sequence 1, Appli
6	27	5.4	14520	15	US-10-156-761-2885 Sequence 2885, Ap
7	27	5.4	15872	9	US-09-861-289-1 Sequence 1, Appli
8	27	5.4	15872	9	US-09-860-846-1 Sequence 1, Appli
9	27	5.4	15872	10	US-09-988-384B-1 Sequence 1, Appli
10	27	5.4	15872	10	US-09-835-821-1 Sequence 1, Appli
11	27	5.4	15872	15	US-10-271-889-44 Sequence 44, Appl
12	27	5.4	125746	15	US-10-156-761-15102 Sequence 15102, A
13	26	5.2	10692	15	US-10-156-761-414 Sequence 414, Appl
14	26	5.2	64492	13	US-10-378-083-1 Sequence 1, Appli

15	5.0	50937	10	US-09-808-880-1	Sequence 1, Appli
16	4.8	3978	10	US-09-953-348-19	Sequence 19, Appl
17	4.8	3978	15	US-10-267-255-19	Sequence 19, Appl
18	4.8	5505	15	US-10-156-761-413	Sequence 413, Appl
19	4.8	5721	15	US-10-156-761-2880	Sequence 2880, Ap
20	4.8	11910	15	US-10-156-761-2879	Sequence 2879, Ap
21	4.8	12249	15	US-09-953-348-74	Sequence 74, Appl
22	4.8	12249	15	US-10-267-255-74	Sequence 74, Appl
23	4.8	18331	10	US-09-953-348-96	Sequence 96, Appl
24	4.8	18331	15	US-10-267-255-96	Sequence 96, Appl
25	4.8	86941	17	US-10-461-194-2	Sequence 2, Appli
26	4.8	125746	15	US-10-156-761-15102	Sequence 15102, A
27	4.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
28	4.6	4674	10	US-09-940-316B-26	Sequence 26, Appl
29	4.6	4725	10	US-09-940-316B-24	Sequence 24, Appl
30	4.6	4737	10	US-09-940-316B-30	Sequence 30, Appl
31	4.6	4767	10	US-09-940-316B-28	Sequence 28, Appl
32	4.6	4818	10	US-09-940-316B-32	Sequence 32, Appl
33	4.6	10056	15	US-10-156-761-411	Sequence 411, Appl
34	4.6	14055	15	US-10-156-761-2883	Sequence 2883, Ap
35	4.6	14643	15	US-10-156-761-931	Sequence 931, Appl
36	4.6	18717	15	US-10-156-761-927	Sequence 927, Appl
37	4.6	20922	13	US-09-942-025-14	Sequence 14, Appl
38	4.6	23238	15	US-10-156-761-415	Sequence 415, Appl
39	4.6	30690	17	US-10-204-862A-1	Sequence 1, Appli
40	4.6	31422	17	US-10-204-862A-2	Sequence 2, Appli
41	4.6	67111	13	US-09-942-025-1	Sequence 1, Appli
42	4.6	80161	17	US-10-329-148A-1	Sequence 1, Appli
43	4.6	85692	17	US-10-461-194-1	Sequence 1, Appli
44	4.6	27705	13	US-09-942-025-12	Sequence 12, Appl
45	4.2	403	15	US-10-259-678-391	Sequence 391, Appl

## ALIGNMENTS

RESULT 1  
US-10-229-148B-1  
; Sequence 1, Application US/10229148B  
; Publication No. US20040091975A1  
; GENERAL INFORMATION:  
; APPLICANT: Meiji Seika Kaisha, Ltd.  
; TITLE OF INVENTION: Midecamycin biosynthetic genes  
; FILE REFERENCE: 138451 US  
; CURRENT APPLICATION NUMBER: US/10/229,148B  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 210516/2002  
; PRIOR FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 84428  
; TYPE: DNA  
; ORGANISM: Streptomyces mycarofaciens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement((1)..(675))  
; OTHER INFORMATION: ORF42 (fragment)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement((1168)..(2202))  
; OTHER INFORMATION: ORF41  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement((2220)..(3215))  
; OTHER INFORMATION: ORF40  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement((3237)..(4691))  
; OTHER INFORMATION: ORF39  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement((4695)..(5948))

```
OTHER INFORMATION: ORF38
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((6048)..(6629))
OTHER INFORMATION: ORF37
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((6653)..(7945))
OTHER INFORMATION: ORF36
FEATURE:
NAME/KEY: CDS
LOCATION: (8149)..(9015)
OTHER INFORMATION: ORF35
FEATURE:
NAME/KEY: CDS
LOCATION: (9012)..(9335)
OTHER INFORMATION: ORF34
FEATURE:
NAME/KEY: CDS
LOCATION: (9328)..(10458)
OTHER INFORMATION: ORF33
FEATURE:
NAME/KEY: CDS
LOCATION: (10521)..(11603)
OTHER INFORMATION: ORF32
FEATURE:
NAME/KEY: CDS
LOCATION: (11729)..(12961)
OTHER INFORMATION: ORF31
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((13016)..(14044))
OTHER INFORMATION: ORF30
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((14074)..(15096))
OTHER INFORMATION: ORF29
FEATURE:
NAME/KEY: CDS
LOCATION: (15643)..(17466)
OTHER INFORMATION: ORF28
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((17522)..(18895))
OTHER INFORMATION: ORF27
FEATURE:
NAME/KEY: CDS
LOCATION: (19063)..(20229)
OTHER INFORMATION: ORF26
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((20307)..(21743))
OTHER INFORMATION: ORF25
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((21733)..(22527))
OTHER INFORMATION: ORF24
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((22534)..(23571))
OTHER INFORMATION: ORF23
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((23555)..(24463))
OTHER INFORMATION: ORF22
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((24460)..(25650))
OTHER INFORMATION: ORF21
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((25647)..(26105))
OTHER INFORMATION: ORF20
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((26180)..(27391))
OTHER INFORMATION: ORF19
FEATURE:
NAME/KEY: CDS
LOCATION: (27937)..(28983)
OTHER INFORMATION: ORF18
FEATURE:
NAME/KEY: CDS
LOCATION: (29244)..(42779)
OTHER INFORMATION: ORF1
FEATURE:
NAME/KEY: CDS
LOCATION: (42823)..(48657)
OTHER INFORMATION: ORF2
FEATURE:
NAME/KEY: CDS
LOCATION: (48712)..(59802)
OTHER INFORMATION: ORF3
FEATURE:
NAME/KEY: CDS
LOCATION: (59850)..(64556)
OTHER INFORMATION: ORF4
FEATURE:
NAME/KEY: CDS
LOCATION: (64687)..(70365)
OTHER INFORMATION: ORF5
FEATURE:
NAME/KEY: CDS
LOCATION: (70365)..(71078)
OTHER INFORMATION: ORF6
FEATURE:
NAME/KEY: CDS
LOCATION: (71113)..(72360)
OTHER INFORMATION: ORF7
FEATURE:
NAME/KEY: CDS
LOCATION: (72400)..(73665)
OTHER INFORMATION: ORF8
FEATURE:
NAME/KEY: CDS
LOCATION: (73694)..(75043)
OTHER INFORMATION: ORF9
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((75899)..(76570))
OTHER INFORMATION: ORF10
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((76602)..(77765))
OTHER INFORMATION: ORF11
FEATURE:
NAME/KEY: CDS
LOCATION: (78039)..(79313)
OTHER INFORMATION: ORF12
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((79391)..(81052))
OTHER INFORMATION: ORF13
FEATURE:
NAME/KEY: CDS
LOCATION: (81541)..(82356)
OTHER INFORMATION: ORF14
FEATURE:
NAME/KEY: CDS
LOCATION: (82760)..(83362)
OTHER INFORMATION: ORF15
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((83495)..(84142))
OTHER INFORMATION: ORF16
FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match          6.4%; Score 32; DB 17; Length 84428;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 GCCCCGACGCGCGTCAAGCGGTGCGCGA 167
Db 60600 GCCCCGACGCGCGTCAAGCGGTGCGCGA 60631

RESULT 2
US-10-156-761-412
; Sequence 412, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 412
; LENGTH: 18435
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(18435)
US-10-156-761-412

Query Match          5.8%; Score 29; DB 15; Length 18435;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 GACGTCGACGCGGTGAGCGCACGCGAC 410
Db 10033 GACGTCGACGCGGTGAGCGCACGCGAC 10061

RESULT 3
US-09-940-316B-1/c
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, INC.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYPEPTIDES ENCODING THE FKBP GENE OF THE FK-520 POLYPEPTIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
```

```
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

Query Match          5.8%; Score 29; DB 10; Length 77536;
Best Local Similarity 100.0%; Pred. No. 5.4e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GCGGTCAACTCCGACGCGCGTCAAGCGG 305
Db 37516 GCGGTCAACTCCGACGCGCGTCAAGCGG 37488

RESULT 4
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Query Match          5.8%; Score 29; DB 15; Length 100000;
Best Local Similarity 100.0%; Pred. No. 5.1e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 GACGTCGACGCGGTGAGCGCACGCGAC 410
Db 60556 GACGTCGACGCGGTGAGCGCACGCGAC 60584

RESULT 5
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```



FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 5.4%; Score 29; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 GACGTGACGGGTGGAGGCGCAGGCAC 410  
|||  
Db 517356 GACGTGACGGGTGGAGGCGCAGGCAC 517328

## RESULT 6

US-10-156-761-2885  
Sequence 2885, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 2885  
LENGTH: 14520  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(14520)

US-10-156-761-2885

Query Match 5.4%; Score 27; DB 15; Length 14520;  
Best Local Similarity 100.0%; Pred. No. 0.00078;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGGCGGTGCAAGGCGTTCGCGG 166  
|||  
Db 5042 CCGACGGCGGTGCAAGGCGTTCGCGG 5068

## RESULT 7

US-09-861-289-1  
Sequence 1, Application US/09861289  
Patent No. US20020110897A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.

APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/861,289  
CURRENT FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: 09/105,537  
PRIOR FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 15872  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-861-289-1

Query Match 5.4%; Score 27; DB 9; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.00076;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131  
|||  
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

## RESULT 8

US-09-860-846-1  
Sequence 1, Application US/09860846  
Patent No. US20020164742A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/860,846  
CURRENT FILING DATE: 2001-05-18  
PRIOR APPLICATION NUMBER: 09/105,537  
PRIOR FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 15872  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-860-846-1

Query Match 5.4%; Score 27; DB 9; Length 15872;  
Best Local Similarity 100.0%; Pred. No. 0.00076;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131  
|||  
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

## RESULT 9

US-09-988-384B-1  
Sequence 1, Application US/09988384B  
Publication No. US20030073824A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.536US1  
CURRENT APPLICATION NUMBER: US/09/988,384B  
CURRENT FILING DATE: 2001-11-19  
PRIOR APPLICATION NUMBER: PCT/US99/14398  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: US 09/105,537

```
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Query Match          5.4%; Score 27; DB 10; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

RESULT 10
US-09-836-821-1
; Sequence 44, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-1

Query Match          5.4%; Score 27; DB 10; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

RESULT 11
US-10-271-889-44
; Sequence 44, Application US/10271889
; Publication No. US20030194784A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
; FILE REFERENCE: 600.582US1
; CURRENT APPLICATION NUMBER: US/10/271,889
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44

; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-271-889-44

Query Match          5.4%; Score 27; DB 15; Length 15872;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 GTTCGTCGAGTTCTCGCGGCGGCGG 131
Db 3721 GTTCGTCGAGTTCTCGCGGCGGCGG 3747

RESULT 12
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match          5.4%; Score 27; DB 15; Length 125746;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 140 CCGACGGCGGCTGCAAGCGGTTTCGCGG 166
Db 48818 CCGACGGCGGCTGCAAGCGGTTTCGCGG 48844

RESULT 13
US-10-156-761-414
; Sequence 414, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 414
; LENGTH: 10692
```

```
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(10692)
US-10-156-761-414

Query Match          5.2%; Score 26; DB 15; Length 10692;
Best Local Similarity 100.0%; Pred.No. 0.0026;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 GTCGACGCGTGGAGGCGCACGCGCAC 410
Db 1003 GTCGACGCGTGGAGGCGCACGCGCAC 1028

RESULT 14
US-10-378-083-1
; Sequence 1, Application US/10378083
; Publication No. US20040053274A1
; GENERAL INFORMATION:
; APPLICANT: President of Tokyo Institute of Technology
; TITLE OF INVENTION: A gene cluster of vicenistatin biosynthase, a vicenistamine
; TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the
; TITLE OF INVENTION: polypeptide
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/378,083
; CURRENT FILING DATE: 2003-03-04
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 64492
; TYPE: DNA
; ORGANISM: Streptomyces halstedii sp.HC-34
US-10-378-083-1

Query Match          5.2%; Score 26; DB 13; Length 64492;
Best Local Similarity 100.0%; Pred.No. 0.0017;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 AGGGCTGCTGGCCACCTACGGCCAG 468
Db 54630 AGGGCTGCTGGCCACCTACGGCCAG 54655

RESULT 15
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-808-880-1
```

```
Query Match          5.0%; Score 25; DB 10; Length 50937;
Best Local Similarity 100.0%; Pred.No. 0.0058;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 GACGGCGCTGCAAGGCGTTCGCGG 166
Db 9582 GACGGCGCTGCAAGGCGTTCGCGG 9606

RESULT 16
US-09-953-348-19
; Sequence 19, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 3978
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-19
```

```
Query Match          4.8%; Score 24; DB 10; Length 3978;
Best Local Similarity 100.0%; Pred.No. 0.032;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 CTGGCCCCCGACGCGCGCTGCAAG 156
Db 757 CTGGCCCCCGACGCGCGCTGCAAG 780

RESULT 17
US-10-267-255-19
; Sequence 19, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 3978
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
```

## US-10-267-255-19

Query Match 4.8%; Score 24; DB 15; Length 3978;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCGAGCGCGCTGCAAG 156  
|||||  
DB 757 CTGGCCCCGAGCGCGCTGCAAG 780  
|||||

## RESULT 18

US-10-156-761-413  
; Sequence 413, Application US/10156761  
; Publication No. US20030119018A1

## GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 413  
; LENGTH: 5505

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(5505)

US-10-156-761-413

Query Match 4.8%; Score 24; DB 15; Length 5505;  
Best Local Similarity 100.0%; Pred. No. 0.03;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 CGCGAGTGCTCCATGGCGCTCGC 71  
|||||  
DB 663 CGCGAGTGCTCCATGGCGCTCGC 686  
|||||

## RESULT 19

US-10-156-761-2880  
; Sequence 2880, Application US/10156761  
; Publication No. US20030119018A1

## GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2880  
; LENGTH: 5721

; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(5721)

US-10-156-761-2880

Query Match 4.8%; Score 24; DB 15; Length 5721;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 GCGCTGCTGGCCACCTACGGCCAG 468  
|||||  
DB 1063 GCGCTGCTGGCCACCTACGGCCAG 1086  
|||||

## RESULT 20

US-10-156-761-2879  
; Sequence 2879, Application US/10156761  
; Publication No. US20030119018A1

## GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2879  
; LENGTH: 11910

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(11910)

US-10-156-761-2879

Query Match 4.8%; Score 24; DB 15; Length 11910;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGCGCTGCTGGCCACCTACGGCC 466  
|||||  
DB 1022 AGCGCTGCTGGCCACCTACGGCC 1045  
|||||

## RESULT 21

US-09-953-348-74  
; Sequence 74, Application US/09953348  
; Publication No. US20030134398A1

## GENERAL INFORMATION:

; APPLICANT: Sherman, David. H  
; APPLICANT: Mao, Yingqing  
; APPLICANT: Varoglu, Mustafa  
; APPLICANT: He, Min  
; APPLICANT: Sheldon, Paul  
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
; FILE REFERENCE: 600.5300S1  
; CURRENT APPLICATION NUMBER: US/09/953,348  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: PCT/US00/06394  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/266965  
; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-74

Query Match 4.8%; Score 24; DB 10; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCCGACGGCGCGCTGCAAG 1767

## RESULT 22

US-10-267-255-74  
; Sequence 74, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-74

Query Match 4.8%; Score 24; DB 15; Length 12249;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCCGACGGCGCGCTGCAAG 1767

## RESULT 23

US-09-953-348-96  
; Sequence 96, Application US/09953348  
; Publication No. US20030134398A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, David. H  
; APPLICANT: Mao, Yingqing  
; APPLICANT: Varoglu, Mustafa  
; APPLICANT: He, Min  
; APPLICANT: Sheldon, Paul  
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
; FILE REFERENCE: 600.530US1  
; CURRENT APPLICATION NUMBER: US/09/953,348  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: PCT/US00/06394  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/266965

; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 96  
; LENGTH: 18331  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-96

Query Match 4.8%; Score 24; DB 10; Length 18331;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCCGACGGCGCGCTGCAAG 1767

## RESULT 24

US-10-267-255-96  
; Sequence 96, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; CURRENT FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 96  
; LENGTH: 18331  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-96

Query Match 4.8%; Score 24; DB 15; Length 18331;  
Best Local Similarity 100.0%; Pred. No. 0.023;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 CTGGCCCCCGACGGCGCGCTGCAAG 156  
|||  
Db 1744 CTGGCCCCCGACGGCGCGCTGCAAG 1767

## RESULT 25

US-10-461-194-2  
; Sequence 2, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, Richard C.  
; APPLICANT: Reid, Ralph C.  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Rascher, Andreas  
; APPLICANT: Schirmer, Andreas  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
; TITLE OF INVENTION: PRO-CELLANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
; FILE REFERENCE: 300622009700  
; CURRENT APPLICATION NUMBER: US/10/461,194

; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: US 60/395,275  
; PRIOR FILING DATE: 2002-07-12  
; PRIOR APPLICATION NUMBER: US 10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/415,326  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US 60/420,820  
; PRIOR FILING DATE: 2002-10-24  
; PRIOR APPLICATION NUMBER: US 60/433,130  
; PRIOR FILING DATE: 2002-12-13  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 86941  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-2

Query Match 4.8%; Score 24; DB 17; Length 86941;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 GCGCTGCTGGCCACCTACGGCCAG 468  
DB 54584 GCGCTGCTGGCCACCTACGGCCAG 54607  
|||||

RESULT 26  
US-10-156-761-15102/c  
; Sequence 15102, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 15102  
; LENGTH: 125746  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-15102

Query Match 4.8%; Score 24; DB 15; Length 125746;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGCGCTGCTGGCCACCTACGGCC 466  
DB 113871 AGCGCTGCTGGCCACCTACGGCC 113848  
|||||

RESULT 27  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.8%; Score 24; DB 15; Length 9025608;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 AGCGCTGCTGGCCACCTACGGCC 466  
DB 3537788 AGCGCTGCTGGCCACCTACGGCC 3537811  
|||||

RESULT 28  
US-09-940-316B-26  
; Sequence 26, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTHETASE  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 4674  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; NAME/KEY: CDS  
; LOCATION: (3)..(4673)  
US-09-940-316B-26

Query Match 4.6%; Score 23; DB 10; Length 4674;  
Best Local Similarity 100.0%; Pred. No. 0.097;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 29

US-09-940-316B-24  
; Sequence 24, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 4725

TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
OTHER INFORMATION: PKS synthase fragment

NAME/KEY: CDS  
LOCATION: (3)..(4724)  
US-09-940-316B-24

Query Match 4.6%; Score 23; DB 10; Length 4725;  
Best Local Similarity 100.0%; Pred. No. 0.097;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 30

US-09-940-316B-30  
; Sequence 30, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551

;; PRIOR FILING DATE: 1999-10-01  
;; PRIOR APPLICATION NUMBER: US 60/139,650  
;; PRIOR FILING DATE: 1999-06-17  
;; PRIOR APPLICATION NUMBER: US 60/123,810  
;; PRIOR FILING DATE: 1999-03-11  
;; PRIOR APPLICATION NUMBER: US 60/102,748  
;; PRIOR FILING DATE: 1998-10-02  
;; NUMBER OF SEQ ID NOS: 72  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 30  
;; LENGTH: 4737

;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:

;; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
;; OTHER INFORMATION: PKS synthase fragment

;; NAME/KEY: CDS  
;; LOCATION: (3)..(4736)  
US-09-940-316B-30

Query Match 4.6%; Score 23; DB 10; Length 4737;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||  
DB 1254 AACTCCGACGGCGGTGCAACGG 1276

## RESULT 31

US-09-940-316B-28  
; Sequence 28, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTU, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH  
; TITLE OF INVENTION: GENE CLUSTER  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 4767

TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
OTHER INFORMATION: PKS synthase fragment

NAME/KEY: CDS  
LOCATION: (3)..(4766)  
US-09-940-316B-28

Query Match 4.6%; Score 23; DB 10; Length 4767;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGGTGCAACGG 305  
|||||

Db 1254 AACTCCGACGGCGTCGAACGG 1276  
|||||

## RESULT 32

US-09-940-316B-32  
; Sequence 32, Application US/09940316B  
; Publication No. US20030175901A1  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTHETASE  
; FILE REFERENCE: 30062-20026.11  
; CURRENT APPLICATION NUMBER: US/09/940,316B  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/410,551  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 4818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; NAME/KEY: CDS  
; LOCATION: (3)..(4817)  
US-09-940-316B-32

Query Match 4.6%; Score 23; DB 10; Length 4818;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 AACTCCGACGGCGTCGAACGG 305  
Db 1254 AACTCCGACGGCGTCGAACGG 1276  
|||||

## RESULT 33

US-10-156-761-411  
; Sequence 411, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 411

; LENGTH: 10056  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(10056)  
US-10-156-761-411

Query Match 4.6%; Score 23; DB 15; Length 10056;  
Best Local Similarity 100.0%; Pred. No. 0.081;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 CCGACGGCGCTGCAAGGGGTC 162  
Db 5237 CCGACGGCGCTGCAAGGGGTC 5259  
|||||

## RESULT 34

US-10-156-761-2883  
; Sequence 2883, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2883  
; LENGTH: 14055  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(14055)  
US-10-156-761-2883

Query Match 4.6%; Score 23; DB 15; Length 14055;  
Best Local Similarity 100.0%; Pred. No. 0.076;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 CTGGGCGACCGATCGAGGCCGA 443  
Db 1012 CTGGGCGACCGATCGAGGCCGA 1034  
|||||

## RESULT 35

US-10-156-761-931  
; Sequence 931, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089



; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 931  
; LENGTH: 14643  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; NAME/KEY: CDS  
; LOCATION: (1)..(14643)  
US-10-156-761-931

Query Match 4.6%; Score 23; DB 15; Length 14643;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 TGGTGGGCGGCTGCCGAGGCT 234  
|||||  
Db 5501 TGGTGGGCGGCTGCCGAGGCT 5523

RESULT 36  
US-10-156-761-927  
; Sequence 927, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 927  
; LENGTH: 18717  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(18717)  
US-10-156-761-927

Query Match 4.6%; Score 23; DB 15; Length 18717;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 ACCCTGGGCGACCCGATCGAGGC 440  
|||||  
Db 1045 ACCCTGGGCGACCCGATCGAGGC 1067

RESULT 37  
US-09-942-025-14  
; Sequence 14, Application US/09942025  
; Publication No. US20030054547A1  
; GENERAL INFORMATION:  
; APPLICANT: Kosan Biosciences, Inc.  
; APPLICANT: Julien, Bryan  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM  
; FILE REFERENCE: 30062-20020.21  
; CURRENT APPLICATION NUMBER: US/09/942,025  
; CURRENT FILING DATE: 2001-08-28

; PRIOR APPLICATION NUMBER: US 60/271,245  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: US 09/144,085  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: US 09/010,809  
; PRIOR FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 20922  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-942-025-14

Query Match 4.6%; Score 23; DB 13; Length 20922;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 CCTGGGCGACCCGATCGAGGCG 442  
|||||  
Db 1032 CCTGGGCGACCCGATCGAGGCG 1054

RESULT 38  
US-10-156-761-415  
; Sequence 415, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 415  
; LENGTH: 23238  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(23238)  
US-10-156-761-415

Query Match 4.6%; Score 23; DB 15; Length 23238;  
Best Local Similarity 100.0%; Pred. No. 0.068;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 GACGGCGCTGCAAGCGTTGCG 164  
|||||  
Db 3796 GACGGCGCTGCAAGCGTTGCG 3818

RESULT 39  
US-10-204-862A-1  
; Sequence 1, Application US/10204862A  
; Publication No. US20040101936A1  
; GENERAL INFORMATION:  
; APPLICANT: ENDO, HIROFUMI  
; APPLICANT: YAMAGUCHI, HIROYUKI  
; APPLICANT: KANDA, YUTAKA  
; APPLICANT: HASHIMOTO, SHINICHI  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

## ; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE

; FILE REFERENCE: 468-32  
; CURRENT APPLICATION NUMBER: US/10/204,862A  
; CURRENT FILING DATE: 2003-12-26  
; PRIOR APPLICATION NUMBER: JP 00/047405  
; PRIOR FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 30690  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(11916)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (11971)..(30687)  
US-10-204-862A-1

Query Match 4.6%; Score 23; DB 17; Length 30690;  
Best Local Similarity 100.0%; Pred. No. 0.063; 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0;

QY 418 ACCCTGGGCGACCCGATCGAGGC 440  
|||||  
Db 13015 ACCCTGGGCGACCCGATCGAGGC 13037  
|||||

## RESULT 40

US-10-204-862A-2  
; Sequence 2, Application US/10204862A  
; Publication No. US20040101936A1  
; GENERAL INFORMATION:  
; APPLICANT: ENDO, HIROFUMI  
; APPLICANT: YAMAGUCHI, HIROYUKI  
; APPLICANT: KANDA, YUTAKA  
; APPLICANT: HASHIMOTO, SHINICHI  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE  
; FILE REFERENCE: 468-32  
; CURRENT APPLICATION NUMBER: US/10/204,862A  
; CURRENT FILING DATE: 2003-12-26  
; PRIOR APPLICATION NUMBER: JP 00/047405  
; PRIOR FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 31422  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(14643)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (14824)..(31419)  
US-10-204-862A-2

Query Match 4.6%; Score 23; DB 17; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 0.063;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 TGGTGGAGCGGCTGTCCGAGGCT 234  
|||||  
Db 5501 TGGTGGAGCGGCTGTCCGAGGCT 5523  
|||||

## RESULT 41

US-09-942-025-1  
; Sequence 1, Application US/09942025  
; Publication No. US20030054547A1

## ; GENERAL INFORMATION:

; APPLICANT: Kosan Biosciences, Inc.  
; APPLICANT: Julien, Bryan  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM  
; TITLE OF INVENTION: CELLULOSUM  
; FILE REFERENCE: 30062-20020.21  
; CURRENT APPLICATION NUMBER: US/09/942,025  
; CURRENT FILING DATE: 2001-08-28  
; PRIOR APPLICATION NUMBER: US 60/271,245  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: US 09/144,085  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: US 09/010,809  
; PRIOR FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 67311  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-942-025-1

Query Match 4.6%; Score 23; DB 13; Length 67311;  
Best Local Similarity 100.0%; Pred. No. 0.053; 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0;

QY 420 CCTGGGCGACCCGATCGAGGCG 442  
|||||  
Db 32581 CCTGGGCGACCCGATCGAGGCG 32603  
|||||

## RESULT 42

US-10-329-148A-1  
; Sequence 1, Application US/10329148A  
; Publication No. US20040023343A1  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/10/329,148A  
; CURRENT FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: US/09/603,207B  
; PRIOR FILING DATE: EARLIER FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 09/370,700  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-10-329-148A-1

Query Match 4.6%; Score 23; DB 17; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0.051;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GACGGCACCGGCTGGTCCGAGGG 197  
|||||  
Db 24771 GACGGCACCGGCTGGTCCGAGGG 24793  
|||||

## RESULT 43

US-10-461-194-1  
; Sequence 1, Application US/10461194  
; Publication No. US20040077058A1  
; GENERAL INFORMATION:

APPLICANT: Hutchinson, Richard C.  
APPLICANT: Reid, Ralph C.  
APPLICANT: Hu, Zhihao  
APPLICANT: Rascher, Andreas  
APPLICANT: Schirmer, Andreas  
APPLICANT: McDaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
PRO-GERANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF  
FILE REFERENCE: 300622009700  
CURRENT APPLICATION NUMBER: US/10/461,194  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130  
PRIOR FILING DATE: 2002-12-13  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 85692  
TYPE: DNA  
ORGANISM: Streptomyces hygroscopicus  
US-10-461-194-1

Query Match 4.6%; Score 23; DB 17; Length 85692;  
Best Local Similarity 100.0%; Pred. No. 0.05;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 CTGGCCACTAGCGCAGAACCG 473  
DB 49924 CTGGCCACTAGCGCAGAACCG 49946

RESULT 44  
US-09-942-025-12  
Sequence 12, Application US/09942025  
Publication No. US20030054547A1  
GENERAL INFORMATION:  
APPLICANT: Kosan Biosciences, Inc.  
APPLICANT: Julien, Bryan  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENE FROM SORANGIUM  
CELLULOSUM  
FILE REFERENCE: 30062-20020.21  
CURRENT APPLICATION NUMBER: US/09/942,025  
CURRENT FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: US 60/271,245  
PRIOR FILING DATE: 2001-02-15  
PRIOR APPLICATION NUMBER: US 09/144,085  
PRIOR FILING DATE: 1998-08-31  
PRIOR APPLICATION NUMBER: US 09/010,809  
PRIOR FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 27705  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-942-025-12

Query Match 4.4%; Score 22; DB 13; Length 27705;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 TCGACGGCGGTGGAGCGCACGG 407  
DB 13082 TCGACGGCGGTGGAGCGCACGG 13103

RESULT 45  
US-10-259-678-391/c  
Sequence 391, Application US/10259678  
Publication No. US20030198974A1  
GENERAL INFORMATION:  
APPLICANT: Cole, Stewart  
APPLICANT: Buchrieser-Brosch, Roland  
APPLICANT: Gordon, Stephen  
APPLICANT: Billault, Alain  
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
FILE REFERENCE: 3495-0169  
CURRENT APPLICATION NUMBER: US/10/259,678  
CURRENT FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US/09/060,756  
PRIOR FILING DATE: 1998-04-16  
NUMBER OF SEQ ID NOS: 743  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 391  
LENGTH: 403  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (various positions within the sequence)  
OTHER INFORMATION: applicants are uncertain of bases designated as "n"  
US-10-259-678-391

Query Match 4.2%; Score 21; DB 15; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 GGTGGAGCGCGCACGGCACCGG 413  
DB 383 GGTGGAGCGCGCACGGCACCGG 363

Search completed: June 16, 2004, 20:21:25  
Job time : 1342 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_8434\_8934

Perfect score: 501

Sequence: 1 tcgtcttcgctggcgcgct.....cgctgtggtcggtcggtg 501

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15

Total number of hits satisfying chosen parameters: 15169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_estc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_est3.\*

12: gb\_est4.\*

13: gb\_est5.\*

14: gb\_estfun.\*

15: em\_estcom.\*

16: em\_esthum.\*

17: em\_estinv.\*

18: em\_estpln.\*

19: em\_estvrt.\*

20: em\_estfun.\*

21: em\_estmam.\*

22: em\_estmus.\*

23: em\_estpro.\*

24: em\_estrod.\*

25: em\_estphg.\*

26: em\_estvrl.\*

27: gb\_est1.\*

28: gb\_est2.\*

29: gb\_est3.\*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	22	4.4	361	13	BY090167
2	22	4.4	574	12	BM397194
3	21	4.2	159	9	AA014791
4	21	4.2	247	14	W65007

5	21	4.2	253	9	AA509737
6	21	4.2	264	13	BY074439
7	21	4.2	309	14	W41546
8	21	4.2	322	14	W41172
9	21	4.2	325	9	AA139922
10	21	4.2	333	13	BY055063
11	21	4.2	335	13	BY052633
12	21	4.2	337	9	AA052685
13	21	4.2	338	10	BB869636
14	21	4.2	345	10	BE859603
15	21	4.2	366	13	BY168519
16	21	4.2	371	13	BY078228
17	21	4.2	372	10	BB871048
18	21	4.2	384	14	W48899
19	21	4.2	396	9	AA914086
20	21	4.2	419	9	AA051350
21	21	4.2	421	13	BY227782
22	21	4.2	448	9	AA020238
23	21	4.2	489	14	CA600107
24	21	4.2	501	9	AA790556
25	21	4.2	544	10	BF581564
26	21	4.2	548	9	AA711986
27	21	4.2	565	12	BM070219
28	21	4.2	573	13	BU582780
29	21	4.2	591	9	AA265232
30	21	4.2	627	29	CE288123
31	21	4.2	644	10	BE286789
32	21	4.2	670	13	BY070379
33	21	4.2	673	13	BU575970
34	21	4.2	686	11	AK002881
35	21	4.2	687	13	BY737644
36	21	4.2	698	10	BE371234
37	21	4.2	702	13	BY709009
38	21	4.2	714	13	BY710477
39	21	4.2	715	11	AK011122
40	21	4.2	728	12	AK009191
41	21	4.2	733	12	BI687681
42	21	4.2	747	10	BE282094
43	21	4.2	761	12	BG976291
44	21	4.2	826	13	BU962156
45	21	4.2	854	14	CB574912

ALIGNMENTS

RESULT 1  
BY090167  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BY090167 361 bp mRNA linear EST 07-DEC-2002  
BY090167 RIKEN full-length enriched, pooled tissues, adult spleen,  
etc. Mus musculus cDNA clone K63080E08 5', mRNA sequence.

BY090167 GI:26199455

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 361)

Okazaki, F., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,

Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bruscia, V.,

Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, I. A.,

Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,

Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,

Kurochkin, I. V., Lee, X., Lenhard, B., Lyons, P. A., Maglott, D. R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,

Petrovsky N., Pillai, R., Pontius J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reid, J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L.G., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Harai, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

further details.

Location/Qualifiers

1. .361

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone\_lib="RIKEN full-length enriched, pooled tissues,

adult spleen, etc."

/note="(dev stage=adult, tissue type=spleen, sex=male),

(dev stage=adult, tissue type=kidney, sex=male),

(dev stage=adult, tissue type=testis, sex=male),

(dev stage=adult, tissue type=thymus, sex=male),

(dev stage=adult, tissue type=heart, sex=male),

(dev stage=adult, tissue type=colon, sex=male),

(dev stage=adult, tissue type=stomach, sex=male),

(dev stage=adult, tissue type=liver, sex=male),

(dev stage=13 days embryo, tissue type=whole

body, sex=mix), (dev stage=14 days embryo, tissue type=whole

body, sex=mix), (dev stage=16 days embryo, tissue type=whole

body, sex=mix), (dev stage=17 days embryo, tissue type=whole

body, sex=mix), (dev stage=15 days pregnant,

adult, tissue type=amnion, sex=female), (dev stage=10 days

# ORIGIN

Query Match 4.4%; Score 22; DB 13; Length 361;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 CCTGGTGGAGCGGCTGTCCGAG 231

DB 64 CCTGGTGGAGCGGCTGTCCGAG 85

## RESULT 2

BM397194/c

LOCUS

DEFINITION

BM397194.1

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

## FEATURES

source

1. .574

/organism="Tetrahymena thermophila"

/mol\_type="mRNA"

/strain="CU428.1"

/db\_xref="taxon:5911"

/clone\_lib="Chilcoat/Turkewitz cDNA (large fraction)"

/note="Vector: BlueScript2 SK+; Details on library

preparation can be found in Chilcoat and Turkewitz (2001)

Proc. Natl. Acad. Sci USA, 98: 8709-8713."

## ORIGIN

Query Match 4.4%; Score 22; DB 12; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 ATGGCGCTCGCGCGCGCGTGA 82

DB 196 ATGGCGCTCGCGCGCGCGTGA 175

## RESULT 3

AA014791

LOCUS

DEFINITION

AA014791

Accession

Version

Keywords

Source

Organism

Reference

neonate, tissue type=brain, sex=mix), (dev stage=10 days  
neonate, tissue type=thymus, sex=mix), (dev stage=10 days  
neonate, tissue type=heart, sex=mix)"

AA014791 159 bp mRNA linear EST 21-JAN-1997  
Clone IMAGE:442898 5', mRNA sequence.  
AA014791 GI:1475654  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 159)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:268234

Seq primer: mob.REGA+ET

High quality sequence stop: 155.

Location/Qualifiers  
1. 159  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:442898"  
/sex="unknown"  
/tissue\_type="placenta"  
/dev\_stage="adult"  
/lab\_host="DH108"  
/clone\_lib="Soares mouse placenta 4NbMPl3.5 14.5"  
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5]."  
TGTTCACATCTGAGTGGCGGCGGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 4.2%; Score 21; DB 9; Length 159;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 48 CTGGTGGAGCGGCTGTCCGAG 68

RESULT 4  
W65007  
LOCUS me04a11.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
DEFINITION clone IMAGE:386492 5', mRNA sequence.

ACCESSION W65007

VERSION W65007.1 GI:1372649

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 247)

REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

AUTHORS

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:238324

Putative full length read

Seq primer: ETPrimer

High quality sequence stop: 238.

Location/Qualifiers  
1. 247  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:386492"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH108"  
/clone\_lib="Soares mouse embryo NbME13.5 14.5"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5]."  
TGTTCACATCTGAGTGGCGGCGGCGGAAATTTTTTTTTTTTTTTTTTTT  
T 3'; on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 4.2%; Score 21; DB 14; Length 247;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 75 CTGGTGGAGCGGCTGTCCGAG 95

RESULT 5  
AA509737  
LOCUS vH52a09.r1 Soares mammary\_gland NbWMG Mus musculus cDNA clone  
DEFINITION IMAGE:890584 5', mRNA sequence.

ACCESSION AA509737

VERSION AA509737.1 GI:2247591

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 253)

REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

AUTHORS

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:518544

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 239.

Location/Qualifiers

#### FEATURES

source

```
1..253
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:890584"
/sex="male"
/tissue_type="mammary gland"
/dev_stages="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares mammary gland NBMWG"
/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo (dT) primer [5'
TGTACCAATCTGAAGTCGAGCGGCCGGAATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
```

#### ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0;

Qy 211 CTGTTGGAGCGGCTGTCCGAG 231

Db 61 CTGTTGGAGCGGCTGTCCGAG 81

#### RESULT 6

BY074439

LOCUS

DEFINITION  
musculus cDNA clone 1920094L17 5', mRNA sequence.

ACCESSION

BY074439.1

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 264)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schönbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,

Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Bruscia, V.,

Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,

Fletcher, C. P., Forrest, A., Frazer, K. S., Gaasterland, T.,

Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,

Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,

Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,

Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,

Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,

Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,

Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,

Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

#### TITLE

Nature 420, 563-573 (2002)

#### JOURNAL

2354683

#### MEDLINE

1246851

#### PUBMED

COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 12-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp/

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, N., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

#### FEATURES

source

1..264

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="I920094L17"

/tissue\_type="kidney"

/dev\_stage="17 days embryo"

/clone\_lib="RIKEN full-length enriched, 17 days embryo

kidney"

#### ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 264;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGTTGGAGCGGCTGTCCGAG 231

Db 109 CTGTTGGAGCGGCTGTCCGAG 129

RESULT 7

W41546

LOCUS

DEFINITION

IMAGE:351932 5', mRNA sequence.

W41546

309 bp mRNA linear EST 11-SEP-1996

mc50811.r1 Soares mouse p3NNF19.5 Mus musculus cDNA clone

W41546

ACCESSION

```

VERSION      W41546.1  GI:1324820
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 309)
AUTHORS      Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
              Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
              Waterston, R.

TITLE        The WashU-HMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:223732

FEATURES     source
              Location/Qualifiers
                1..309
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="IMAGE:351932"
                /dev_stage="19.5 dpc total fetus"
                /lab_host="DH10B (ampicillin resistant)"
                /clone_lib="Soares mouse p3NNMF19.5"
                /note="Vector: pT7T3D (Pharmacia) with a modified
                polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5',
                TGTTCACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT7T3 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by Bento
                Soares and M.Fatima Bonaldo. RNA was kindly provided by
                Dr. Minoru Ko (Wayne State University)."
```

## ORIGIN

```

Query Match      4.2%; Score 21; DB 14; Length 309;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
    |||||
Db 36 CTGGTGGAGCGGCTGTCCGAG 56

RESULT 8
W41172
LOCUS      mc42g01.r1 Soares mouse p3NNMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:351216 5', mRNA sequence.
ACCESSION  W41172.1  GI:1325507
VERSION     W41172.1  GI:1325507
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 322)
AUTHORS      Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
              Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
              Waterston, R.

TITLE        The WashU-HMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:223732
```

```

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:223016

FEATURES     source
              Location/Qualifiers
                1..322
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="IMAGE:351216"
                /dev_stage="19.5 dpc total fetus"
                /lab_host="DH10B (ampicillin resistant)"
                /clone_lib="Soares mouse p3NNMF19.5"
                /note="Vector: pT7T3D (Pharmacia) with a modified
                polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5',
                TGTTCACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT7T3 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by Bento
                Soares and M.Fatima Bonaldo. RNA was kindly provided by
                Dr. Minoru Ko (Wayne State University)."
```

## ORIGIN

```

Query Match      4.2%; Score 21; DB 14; Length 322;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
    |||||
Db 54 CTGGTGGAGCGGCTGTCCGAG 74

RESULT 9
AA139922
LOCUS      mq37h05.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:580953 5',
DEFINITION mRNA sequence.
ACCESSION  AA139922
VERSION     AA139922.1  GI:1702126
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 325)
AUTHORS      Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
              Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
              Waterston, R.

TITLE        The WashU-HMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:223016
```



Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yaenunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.	
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
Nature 420, 563-573 (2002)	
22354683	
1246851	
Contact: Yoshihide Hayashizaki	
Laboratory for Genome Exploratory Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute	
The Institute of Physical and Chemical Research (RIKEN)	
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan	
Tel: 81-45-503-9222	
Fax: 81-45-503-9216	
Email: genome-res@gs.c.riken.go.jp,	
URL: http://genome.gsc.riken.go.jp/	
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsgami, M., Waki, K., Watanuki, A., Muramatsu, M. and Hayashizaki, Y.	
Direct Submission	
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)	

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

```

FEATURES
source
Location/Qualifiers
1. .333
/organism="Mus musculus"
/mol_type="mRNA"
/strain="DBA/2"
/db_xref="taxon:10090"
/clone="I730080N15"
/cell_line="CRL-1722 L5178Y-R"
/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"

ORIGIN
Query Match 4.2%; Score 21; DB 13; Length 333;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231
|||||
Db 102 CTGGTGGAGCGGCTGTCCGAG 122

RESULT 11
BY052633
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BY052633 335 bp mRNA linear EST 06-DEC-2002
BY052633 RIKEN full-length enriched, TIB-55 B88 Mus musculus cDNA
clone I730070L04 5', mRNA sequence.
BY052633
BY052633.1 GI:26158081
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

||||| 54 CTGGTGGAGCGGCTGTCGAG 74

Db
RESULT 13
BB869636 338 bp mRNA linear EST 27-NOV-2001
LOCUS BB869636 RIKEN full-length enriched, pooled tissues, intestinal
DEFINITION BB869636 Mus musculus cDNA clone G630014K09 5', mRNA sequence.
ACCESSION BB869636
VERSION BB869636.1 GI:17115846
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 338)
Akimura,T., Arawaka,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Iehikawa,T., Ozawa,K., Tanaka,T.,
Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
source
location/Qualifiers
1..338
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="G630014K09"
/clone_lib="RIKEN full-length enriched, pooled tissues,
intestinal mucosa, etc."
/note="pooled tissues ; (tissue type=intestinal mucosa,
dev_stage=adult, sex=male), (tissue type=accessory
axillary lymph node, dev_stage=adult, sex=male),
(tissue type=jejunal and colic lymph node,
dev_stage=adult, sex=male), (tissue type=gall bladder,
dev_stage=adult, sex=male), (tissue_type=vesicular gland,
dev_stage=adult, sex=male)"

dev_stage=adult, sex=male), (tissue_type=spinal cord,
dev_stage=11 days embryo), (tissue_type=brain,
dev_stage=13 days embryo), (tissue_type=epinal cord,
dev_stage=13 days embryo), (tissue_type=lung, dev_stage=14
days embryo), (tissue_type=brain, dev_stage=15 days
embryo), (tissue_type=ovary and uterus, dev_stage=10 days
pregnant adult, sex=female), (tissue_type=cerebellum,
dev_stage=0 day neonate), (tissue_type=diencephalon,
dev_stage=1 month neonate), (tissue_type=diencephalon,
dev_stage=16 days neonate, sex=male), (tissue_type=medulla
oblongata, dev_stage=16 days neonate, sex=male),
(tissue_type=cerebellum, dev_stage=21 days neonate),
(tissue_type=testis, dev_stage=8 days neonate, sex=male)"

ORIGIN
Query Match 4.2%; Score 21; DB 10; Length 338;
Best Local Similarity 100.0%; Pred.No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 211 CTGGTGGAGCGGCTGTCGAG 231
|||||
Db 108 CTGGTGGAGCGGCTGTCGAG 128
|||||

RESULT 14
BB859603 345 bp mRNA linear EST 29-SEP-2000
LOCUS BB859603
DEFINITION UI-M-AP0-abe-c-05-0-UI.r1 NIH_BMAP_MST Mus musculus cDNA clone
ACCESSION BB859603
VERSION BB859603.1 GI:10375692
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 345)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE 97044477
PUBMED 8889548
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
FEATURES
source
location/Qualifiers
1..345
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AP0-abe-c-05-0-UI"
/clone_lib="NIH_BMAP MST"
/lab_host="DHI0B (Life Technologies)"
/dev_stage="27-32 days"
/note="Vector: pMT23D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MST library is a non-normalized library
constructed from mouse striatum. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described

```

by Bonaldo, Lennon and Soares, Genome Research 6:  
791-806, 1996. Tissue provided by Ms. Annie Novakovich,  
Zivic-Wilmer Laboratories."

ORIGIN

Query Match 4.2%; Score 21; DB 10; Length 345;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGCTGGAGCGGCTGTCGAG 231  
Db 89 CTGCTGGAGCGGCTGTCGAG 109

RESULT 15  
BY168519  
LOCUS  
DEFINITION  
Mus musculus full-length enriched, bone marrow macrophage Mus  
musculus cDNA clone I830072B19 5', mRNA sequence.

ACCESSION  
BY168519  
VERSION  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 366)

REFERENCE  
AUTHORS  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Matsuda, H., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, K.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehi, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683  
12466851

CONTACT: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numata, K.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by David A. Hume ( Depts. of Biochemistry  
and Microbiology/Parasitology Institute for Molecular Bioscience  
University of Queensland Brisbane, Q 4072 Australia ) whose  
assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details

# FEATURES

Location/Qualifiers  
1. 366  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I830072B19"  
/tissue\_type="bone marrow"  
/cell\_type="macrophage"  
/clone\_lib="RIKEN full-length enriched, bone marrow  
macrophage"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 366;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGCTGGAGCGGCTGTCGAG 231  
Db 54 CTGCTGGAGCGGCTGTCGAG 74

## RESULT 16

BY078228  
LOCUS  
DEFINITION  
RIKEN full-length enriched, pooled tissues, adult spleen,  
etc. Mus musculus cDNA clone K630013G08 5', mRNA sequence.

ACCESSION  
BY078228  
VERSION  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 371)

## REFERENCE

## AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Matsuda, H., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,

Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Sakazume, N., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

22354683  
 12466951

**TITLE**  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Itoh, M., Kawai, J., Konno, H., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numasaki, R., Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

**FEATURES**  
 source  
 Location/Qualifiers  
 1. 371  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."  
 /note="(dev stage=adult, tissue type=spleen, sex=male), (dev stage=adult, tissue type=kidney, sex=male), (dev stage=adult, tissue type=testis, sex=male), (dev stage=adult, tissue type=thymus, sex=male), (dev stage=adult, tissue type=heart, sex=male), (dev stage=adult, tissue type=colon, sex=male), (dev stage=adult, tissue type=stomach, sex=male), (dev stage=adult, tissue type=livver, sex=male), (dev stage=13 days embryo, tissue type=whole body, sex=mix), (dev stage=14 days embryo, tissue type=whole body, sex=mix), (dev stage=16 days embryo, tissue type=whole body, sex=mix), (dev stage=17 days embryo, tissue type=whole body, sex=mix), (dev stage=15 days pregnant, adult, tissue type=amion, sex=female), (dev stage=10 days neonate, tissue type=brain, sex=mix), (dev stage=10 days

**ORIGIN**

Query Match 4.2%; Score 21; DB 13; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCGAG 231

DB 114 CTGGTGGAGCGGCTGTCGAG 134

**RESULT 17**

BB871048 372 bp mRNA linear EST 27-NOV-2001  
 BB871048 RIKEN full-length enriched, 13 days embryo spinal cord Mus  
 musculus cDNA clone G630026J17 5', mRNA sequence.

BB871048

BB871048.1 GI:17117258

EST.

SOURCE Mus musculus (house mouse)

**ORGANISM**

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE**

1 (Bases 1 to 372)  
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, K., Okido, T., Saito, K., Saito, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

**TITLE****JOURNAL****COMMENT**

Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

**FEATURES****source**

Location/Qualifiers

1. 372

/organism="Mus musculus"

/mol\_type="mRNA"

/strains="CS7BL/6J"

/db\_xref="taxon:10090"



Db 29 CTGGTGGAGCGGCTGTCGAG 49

RESULT 20  
AA051350  
LOCUS  
DEFINITION  
ct52h02.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA  
IMAGE:479763 5', mRNA sequence.

ACCESSION  
AA051350  
VERSION  
AA051350.1 GI:1531217  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 419)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:290507  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 349.  
Location/Qualifiers  
1. .419  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:479763"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NM013.5 14.5"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', TGTTCACCAATCTGAATGGAGCGCGCGCGAATTTTTTTTTTTTTTTTTTTT 3', on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Patina Bonaldo."

ORIGIN  
Query Match 4.2%; Score 21; DB 9; Length 419;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCGAG 231  
|||||  
Db 75 CTGGTGGAGCGGCTGTCGAG 95  
|||||

RESULT 21  
BY227782  
LOCUS

DEFINITION  
BY227782 RIKEN full-length enriched, bone marrow mast cells Mus musculus cDNA clone KIC0003804 5', mRNA sequence.

ACCESSION  
BY227782  
VERSION  
BY227782.1 GI:26408892  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 421)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Hume, D. A., Quackenbush, J., Baldarelli, R., Hill, D. P., Bult, C., Schat, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Athia, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierzki, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, D. J., Reid, J., King, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, B., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.ssc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Hiromitsu Nakauchi ( Dept. of Immunology  
Institute of Basic Medical Sciences University of Tsukuba 1-1-1  
Tennodai,Tsukuba,Ibaraki 305-8578 ) whose assistance we gratefully  
acknowledge. Please visit our web site  
(<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

```
1. 421
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K1C0003B04"
/tissue_type="bone marrow"
/cell_type="mast cells"
/clone_lib="RIKEN full-length enriched, bone marrow mast
cells"
```

## ORIGIN

```
Query Match 4.2%; Score 21; DB 13; Length 421;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 211 CTGTGGAGCGGCTGTCCGAG 231
|||||
DB 63 CTGTGGAGCGGCTGTCCGAG 83
|||||
```

## RESULT 22

AA020238

LOCUS

```
DEFINITION mh49g09.r1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA
clone IMAGE:445888 5', mRNA sequence.
```

ACCESSION

AA020238

VERSION

AA020238.1 GI:1484004

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

REFERENCE

AUTHORS

```
1 (bases 1 to 448)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, P., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
```

TITLE

JOURNAL

COMMENT

```
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:271224
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 439.
```

## FEATURES

source

```
1. 448
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:445888"
/sex="unknown"
/tissue_type="placenta"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: placenta; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
```

```
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTTACCAATCTGAAGTCGGAGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."
```

## ORIGIN

```
Query Match 4.2%; Score 21; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 211 CTGTGGAGCGGCTGTCCGAG 231
|||||
DB 94 CTGTGGAGCGGCTGTCCGAG 114
|||||
```

## RESULT 23

CA600107

LOCUS

```
DEFINITION wawic.pk006.k4 waw1c Triticum aestivum cDNA clone wawic.pk006.k4 5',
end, mRNA sequence.
```

ACCESSION

CA600107

VERSION

CA600107.1

KEYWORDS

EST.

SOURCE

Triticum aestivum (bread wheat)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```
1 (bases 1 to 489)
Tingey, S.V., Langridge, P., Powell, W., Wolters, P., Dolan, M.,
Hainey, C., Miao, G., Caraher, N. and Hanafey, M.K.
DuPont Wheat cDNA Sequence in collaboration with the Waite
Institute
Unpublished (2002)
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
```

FEATURES

source

```
1. 489
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="wawic.pk006.k4"
/tissue_type="anthers"
/lab_host="DH10B"
/clone_lib="wawic"
/notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; To
examine gene expression patterns in wheat anthers
undergoing meiosis. Library was made at the Waite
Institute in Australia."
```

## ORIGIN

```
Query Match 4.2%; Score 21; DB 14; Length 489;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 436 GAGGCCGAGCGGCTGTGGCC 456
|||||
DB 188 GAGGCCGAGCGGCTGTGGCC 208
|||||
```

## RESULT 24

AA790556



LOCUS AA790556 501 bp mRNA linear EST 06-FEB-1998  
 DEFINITION v47h08.r1 Soares mammary\_gland\_NbMMG Mus musculus cDNA clone  
 IMAGE:1244127 5', mRNA sequence.

ACCESSION AA790556  
 VERSION AA790556.1 GI:2850676  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 501)  
 REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:657815  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 500.

FEATURES  
 Location/Qualifiers  
 1..501  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CS7BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1244127"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares mammary\_gland NbMMG"  
 /note="Organ: mammary gland; Vector: pT7T3D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5'  
 TGTTCACCATCTGAAGTGGAGCGCGCGAATGGTTTCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

ORIGIN  
 Query Match 4.2%; Score 21; DB 9; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
 |||||  
 Db 44 CTGGTGGAGCGGCTGTCCGAG 64  
 |||||

RESULT 25  
 BF581564 544 bp mRNA linear EST 12-DEC-2000  
 LOCUS 602101087F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4224572 5',  
 DEFINITION mRNA sequence.

ACCESSION BF581564  
 VERSION BF581564.1 GI:11655276  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 544)  
 REFERENCE NIH-MGC http://mgs.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM9814 row: m column: 21  
 High quality sequence stop: 539.

FEATURES  
 Location/Qualifiers  
 1..544  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4224572"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI CGAP Co24"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.6 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN  
 Query Match 4.2%; Score 21; DB 10; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
 |||||  
 Db 55 CTGGTGGAGCGGCTGTCCGAG 75  
 |||||

RESULT 26  
 AA711986 548 bp mRNA linear EST 24-DEC-1997  
 LOCUS v460a07.r1 Soares mammary\_gland\_NbMMG Mus musculus cDNA clone  
 DEFINITION IMAGE:1195764 5', mRNA sequence.

ACCESSION AA711986  
 VERSION AA711986.1 GI:2721904  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 548)  
 REFERENCE Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:642860  
 Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 506.

## FEATURES

source

1..548  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1195764"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stages="4 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares mammary gland NBMNG"

/note="Vector: pSPORT1; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGCGCAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 4.2%; Score 21; DB 9; Length 548;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCGCGAG 231

Db 69 CTGGTGGAGCGGCTGTCGCGAG 89

## RESULT 27

BM070219

LOCUS

BM070219 id79c06.yi Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus  
musculus cDNA clone IMAGE:5669363 5', mRNA sequence.

## ACCESSION

BM070219

VERSION

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

565 bp mRNA linear EST 12-MAR-2002  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 565)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,  
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R.,  
Williams, T., Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Juliana Brown  
(brownjfas.harvard.edu)  
MGI:1955689 This sequence now available from the IMAGE consortium,  
for clone orders contact: info@image.llnl.gov  
High quality sequence stop: 435.  
Location/Qualifiers  
1..565

## FEATURES

source

/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="ICR"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5669363"  
/sex="Both for embryonic & newborn, male for adult and  
adult islet"  
/dev stage="Embryonic day 10.5, E12.5, E16.5, newborn,  
adult, mixed"  
/lab host="DH10B"  
/clone\_lib="Melton Normalized Mixed Mouse Pancreas 1  
N1-MMS1"  
/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five  
libraries representing E10.5/12.5 pancreatic bud, E16.5  
pancreas, newborn pancreas, adult pancreas, and adult  
islets of Langerhans were separately constructed using  
SuperScript Plasmid Library kit (Life Technologies). cDNA  
was made by oligo-dT priming and size-selected by column  
fractionation. Libraries were amplified once on solid  
support and plasmid DNA from each library was prepared  
and mixed in equal amounts. The mixed library DNA was  
normalized by method #4 from Bonaldo, Lennon, and Soares  
1996 Genome Research 6:791-806; 0.5 microgram  
single-stranded mixed library plasmid DNA was mixed with  
5 micrograms PCR product representing mixed library  
inserts and hybridized to an EcoT of 6. Single-stranded  
(unhybridized) plasmids were isolated by hydroxyapatite  
chromatography and used to make this library."

## ORIGIN

Query Match 4.2%; Score 21; DB 12; Length 565;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCGCGAG 231

Db 84 CTGGTGGAGCGGCTGTCGCGAG 104

## RESULT 28

BU582780

LOCUS

DEFINITION

spematocytes Mus musculus cDNA clone IMAGE:6366860 5', mRNA

sequence.

ACCESSION

BU582780

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

573 bp mRNA linear EST 20-SEP-2002  
mah78fii.y1 McCarrey Eddy 18 day leptotene and zygotene  
sequence.  
BU582780  
BU582780.1 GI:23256745  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 573)  
McCarrey, J., Eddy, M., Marra, M., Hillier, L., Clifton, S., Pape, D.,  
Martin, J., Wylie, T., Dante, M., Bowers, Y., Theising, B., Gibbons, M.,  
Ritter, E., Teagareishvili, R., Ronko, I., Maguire, L., Kennedy, S.,  
Bennett, J., Waterston, R. and Wilson, R.  
NIHES Mouse  
Unpublished (2002)  
Contact: McCarrey/Eddy NIHES Mouse  
NIHES Mouse  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed and donated by J. McCarrey, Ph.D. (Southwest  
Foundation for Biomedical Research, Dept. of Genetics) - excision  
done by E.M. Eddy, Ph.D. (National Institutes of Health, National  
Institute of Environmental Health Sciences).  
MGI:2044292  
Seq primer: -40RP from Gibco  
High quality sequence stop: 423.



```

RESULT 31
LOCUS BE286789
DEFINITION BE286789 644 bp mRNA linear EST 26-OCT-2000
601090518F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3485141 5',
mRNA sequence.
ACCESSION BE286789
VERSION BE286789.1 GI:9165352
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 644)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLaM8519 row: d column: 06
High quality sequence stop: 585.
Location/Qualifiers
1. .644
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3485141"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam5"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

ORIGIN
Query Match 4.2%; Score 21; DB 10; Length 644;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGTGGAGCGGCTGTCCGAG 231
Db 119 CTGTGGAGCGGCTGTCCGAG 139

RESULT 32
LOCUS BY703379
DEFINITION BY703379 670 bp mRNA linear EST 16-DEC-2002
BY703379 RIKEN full-length enriched, adult male kidney Mus musculus
cDNA clone 0610041D15 5', mRNA sequence.
ACCESSION BY703379
VERSION BY703379.1 GI:27114486
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 670)
Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
Nikaido I., Osato N., Saito K., Suzuki H., Yamanaka I.,
Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A.,
Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C.,
Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H.,
Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V.,

```

```

Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Guscincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kenzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
Sutarada, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Komori, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
Location/Qualifiers
1. .670
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="0610041D15"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male kidney"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was

```

## FEATURES

source

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGCGCGCGTGTCTCCGAG 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGAGAGTCCAGAGCTCAATTAATTAATPARACCCGCCCC 3']. cDNA was cleaved with XhoI and SstI."

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 670;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 211 CTGCTGGAGCGCTGTCCGAG 231  
|||||  
Db 55 CTGCTGGAGCGCTGTCCGAG 75  
|||||

## RESULT 33

BU757970/c  
LOCUS 673 bp mRNA linear EST 10-OCT-2002  
DEFINITION UI-1-CF0-apb-d-09-0-UI.s1 NCI CGAP\_P1Tr1 Mus musculus cDNA clone  
UI-1-CF0-apb-d-09-0-UI 3', mRNA sequence.

## ACCESSION

VERSION BU757970.1 GI:23720040

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 673)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

## AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

## JOURNAL

Unpublished (1997)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Minoru Ko, Dr. Janet Rossent, Dr. Janet Rossent  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

Location/Qualifiers

1..673  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-1-CF0-apb-d-09-0-UI"  
/tissue\_type="Trophoblast"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP\_P1Tr1"  
/note="Organ: Placenta; Vector: pTTT3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
NCI CGAP P1Tr1 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The

sequence tags for this library are GTGTG, ATCAT, GGGTG, GGTGTG. For additional information, contact: Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
TAG TISSUE=placenta mouse  
TAG LIB=UI-1-CF0  
TAG SEQ=GTGTG"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 673;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 211 CTGCTGGAGCGCTGTCCGAG 231  
|||||  
Db 650 CTGCTGGAGCGCTGTCCGAG 630  
|||||

## RESULT 34

AK002881

## LOCUS

DEFINITION

Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610041D15 product:hypothetical protein, full insert sequence.

## ACCESSION

VERSION AK002881.1 GI:12833190

## KEYWORDS

HTC; CAP trapper.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1

## AUTHORS

Carninci, P. and Hayashizaki, Y.

## TITLE

High-efficiency full-length cDNA cloning

## JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

## MEDLINE

99279253

## PUBMED

10349636

## REFERENCE

2

## AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

## TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

## JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

## MEDLINE

20499374

## PUBMED

11042159

## REFERENCE

3

## AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

## TITLE

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

## JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

## MEDLINE

20530913

## PUBMED

11076861

## REFERENCE

4

## AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

## TITLE

Functional annotation of a full-length mouse cDNA collection

## JOURNAL

Nature 409, 685-690 (2001)

## PUBMED

11076861

## REFERENCE

5

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

## TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

## JOURNAL

Nature 420, 563-573 (2002)

## PUBMED

12051111

## REFERENCE

6 (bases 1 to 686)

## AUTHORS

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirata, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,



prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details

# FEATURES

source  
Location/Qualifiers  
1. .687  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="I73008M03"  
/cell\_line="TIB-55 BB88"  
/clone\_lib="RIKEN full-length enriched, TIB-55 BB88"

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 687;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGCTGGAGCGGCTGCTCCGAG 231  
|||||  
DB 111 CTGCTGGAGCGGCTGCTCCGAG 131  
|||||

## RESULT 36

BE371234 601218984F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3587780 5', linear EST 21-JUL-2000  
LOCUS BE371234 698 bp mRNA  
DEFINITION BE371234.1 GI:9316597  
VERSION BE371234  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 698)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8751 row: h column: 21  
High quality sequence stop: 626.  
Location/Qualifiers  
1. .698  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3587780"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
Stem cell origin."  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lu29"  
/note="Organ: Lung; Vector: pCMV-SF0RT6; Site 1: SalI;  
Site 2: NotI; cloned unidirectionally. Primer: Oligo dn.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

## FEATURES

source  
Location/Qualifiers  
1. .698  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3587780"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
Stem cell origin."  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lu29"  
/note="Organ: Lung; Vector: pCMV-SF0RT6; Site 1: SalI;  
Site 2: NotI; cloned unidirectionally. Primer: Oligo dn.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

## ORIGIN

Query Match 4.2%; Score 21; DB 10; Length 698;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 211 CTGCTGGAGCGGCTGCTCCGAG 231  
|||||

## Db

RESULT 37  
BY709009  
LOCUS BY709009  
DEFINITION BY709009  
ACCESSION BY709009  
VERSION BY709009  
KEYWORDS BY709009  
SOURCE BY709009  
ORGANISM BY709009

## REFERENCE

### AUTHORS

1 (bases 1 to 702)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wyshew-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
22354683  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@gs.riken.go.jp](mailto:genome-res@gs.riken.go.jp),  
[URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
Sato, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

57 CTGCTGGAGCGGCTGCTCCGAG 77

BY709009 702 bp mRNA linear EST 16-DEC-2002  
BY709009 RIKEN full-length enriched, adult male tongue Mus musculus  
cDNA clone 2310006P18 5', mRNA sequence.

BY709009

BY709009.1 GI:27120205

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 702)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wyshew-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
22354683  
12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@gs.riken.go.jp](mailto:genome-res@gs.riken.go.jp),  
[URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
Sato, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. .702

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="2310006P18"

/sex="male"

/tissue type="tongue"

/dev stage="adult"

/lab\_host="SOLR"

/clone\_lib="RIKEN full-length enriched, adult male tongue"

/note="Site 1: XhoI; Site 2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGTCGAGGCTCTTCGAG 231], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGTCGAGGCTCTTCGAGTCTTCGAG 113]. cDNA was cleaved with XhoI and SstI."

## ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 702;

Best Local Similarity 100.0%; Pred. No. 3.6e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTCTTCGAG 231

Db 93 CTGGTGGAGCGGCTCTTCGAG 113

## RESULT 38

BY710477

LOCUS

DEFINITION BY710477 RIKEN full-length enriched, 13 days embryo liver Mus

musculus cDNA clone 2510049L02 5', mRNA sequence.

ACCESSION

BY710477

VERSION

EST.

BY710477.1

GI:27121700

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 714)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,

Catholia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,

Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,

Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sadelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Varado, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, J.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-3222

Fax: 81-45-503-3216

Email: [genome-res@gs.riken.go.jp](mailto:genome-res@gs.riken.go.jp)

URL: <http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. .714

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="2510049L02"

/tissue type="liver"

/dev stage="13 days embryo"

/lab host="SOLR"

/clone\_lib="RIKEN full-length enriched, 13 days embryo liver"

/note="Site 1: XhoI; Site 2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGTCGAGGCTCTTCGAGTCTTCGAGTCTTCGAG 3']. cDNA was prepared by using trehalose thermo-activated reverse





prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCTTCAGTTAATTAATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.

**FEATURES**  
source  
1. .728  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:2310006P18"  
/db\_xref="MGI:1897547"  
/db\_xref="taxon:10090"  
/clone="2310006P18"  
/sex="male"  
/tissue\_type="tongue"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
<1. .476  
/note="unnamed protein product; hypothetical protein (evidence: Truncated-Longest-ORF)  
putative"  
/codon\_start=3  
/protein\_id="BAB26131.1"  
/db\_xref="GI:12843833"  
/translation="PEGSERRVPMRGESHLGPAAGGPAAPGAAGAAVVAHPA  
GAAHSLRAATAARPCSPAPRPGSAARCPASPREVQGHFGRTPPPGPR  
MAEGSGRSQVSTDTWTLPGRGHVPSVIYTDLSHSSPSGRVAASPNSDCPTDA"  
712. .717  
polyA\_signal  
/note="putative"  
728  
polyA\_site  
/note="putative"  
ORIGIN

Query Match 4.2%; Score 21; DB 11; Length 728;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
|||||  
Db 97 CTGGTGGAGCGGCTGTCCGAG 117  
|||||

RESULT 41  
BI687681 733 bp mRNA linear EST 18-SEP-2001  
LOCUS 603312682F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5352549 5',  
DEFINITION mRNA sequence.  
ACCESSION BI687681  
VERSION BI687681.1 GI:15650309  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 733)  
NTH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgabs-x@mail.nih.gov](mailto:cgabs-x@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM11896 row: d column: 22  
High quality sequence stop: 669.  
Location/Qualifiers  
1. .733

**FEATURES**  
source

AK009191.1 GI:12843832  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
PUBMED 12010728  
6 (bases 1 to 728)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216]  
COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCCAGAGCTCTCTTTTCTTTTCTTTT 3'], cDNA was

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5352549"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 12; Length 733;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCGAG 231
Db 78 CTGGTGGAGCGGCTGTCGAG 98

RESULT 42
BE282094 747 bp mRNA linear EST 26-OCT-2000
LOCUS 601100912F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3493299 5',
DEFINITION mRNA sequence.
ACCESSION BE282094.1 GI:9157425
VERSION BE282094
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 747)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8540 row: h column: 04
High quality sequence stop: 638.
Location/Qualifiers
1..747
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:3493299"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu29"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 10; Length 747;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 211 CTGGTGGAGCGGCTGTCGAG 231
Db 59 CTGGTGGAGCGGCTGTCGAG 79

RESULT 43
BG976291 761 bp mRNA linear EST 12-JUN-2001
LOCUS 602846558F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4976991 5',
DEFINITION mRNA sequence.
ACCESSION BG976291
VERSION BG976291.1 GI:14363928
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 761)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10971 row: d column: 16
High quality sequence stop: 681.
Location/Qualifiers
1..761
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976991"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match          4.2%; Score 21; DB 12; Length 761;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 211 CTGGTGGAGCGGCTGTCGAG 231
Db 88 CTGGTGGAGCGGCTGTCGAG 108

RESULT 44
BU962156 826 bp mRNA linear EST 21-OCT-2002
LOCUS AGENCOURT_10614426 NIH_MGC_169 Mus musculus cDNA clone
DEFINITION IMAGE:6742852 5', mRNA sequence.
ACCESSION BU962156
VERSION BU962156.1 GI:24191740
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 826)
NIH-MGC http://mgc.nci.nih.gov/

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Jonathan Kuo, NIMH  
 CDNA Library Preparation: Michael Brownstein Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LICM3081 row: j column: 03  
 High quality sequence stop: 409.  
 Location/Qualifiers

FEATURES  
 source  
 1..826  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6742852"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_169"  
 /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgccctcgcc); CDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
 5'-RAGCAGTGTATCAGCGAGAGTGGCCATTACGCGGG-3' and  
 5'-ATTCTAGCGCGAGCGGCCACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

# ORIGIN

Query Match 4.2%; Score 21; DB 13; Length 826;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
 DB 49 CTGGTGGAGCGGCTGTCCGAG 69

RESULT 45  
 CB574912 854 bp mRNA linear EST 02-APR-2003  
 LOCUS AGENCOURT\_12974065 NIH\_MGC\_136 Mus musculus CDNA clone  
 DEFINITION IMAGE:30291619 5', mRNA sequence.  
 CB574912  
 CB574912.1 GI:29494442  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 854)  
 NIH-MGC <http://mgi.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. David Rowe  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: NDAM334 row: c column: 20  
 High quality sequence start: 8  
 High quality sequence stop: 622.  
 Location/Qualifiers

FEATURES  
 source  
 1..854  
 /organism="Mus musculus"

# ORIGIN

Query Match 4.2%; Score 21; DB 14; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CTGGTGGAGCGGCTGTCCGAG 231  
 DB 126 CTGGTGGAGCGGCTGTCCGAG 146

Search completed: June 16, 2004, 19:29:20  
 Job time : 2108.33 secs

/mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30291619"  
 /tissue\_type="embryonic limb, maxilla and mandible"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_136"  
 /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:02:16 ; Search time 2074.33 Seconds  
(without alignments)  
10468.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcgtcccgagccgagca.....ccctcgggcagacgagctga 501

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 15  
Total number of hits satisfying chosen parameters: 8604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	53784	1	AMM223012	AJ223012 Amycolato
2	501	100.0	53789	6	A69720	A69720 Sequence 3
3	501	100.0	90445	1	AF040570	AF040570 Amycolato
4	31	6.2	82746	1	AF453501	AF453501 Actinosyn
5	28	5.6	41097	1	AF016585	AF016585 Streptomy
6	26	5.2	6567	1	AF521897	AF521897 Streptomy
7	26	5.2	22131	1	AF506522	AF506522 Streptomy
8	26	5.2	132544	1	AF521085	AF521085 Streptomy
9	24	4.8	78210	1	AB070949	AB070949 Streptomy
10	24	4.8	300425	1	AP005022	AP005022 Streptomy
11	23	4.6	6378	6	AX751873	AX751873 Sequence
12	23	4.6	18614	1	AE007033	AE007033 Mycobacte
13	23	4.6	31422	6	E38021	E38021 Avermectin
14	23	4.6	31422	6	BD097650	BD097650 A method
15	23	4.6	64957	1	AB032367	AB032367 Streptomy
16	23	4.6	132544	1	AF521085	AF521085 Streptomy
17	23	4.6	300050	1	BX248339	BX248339 Mycobacte
18	23	4.6	302675	1	AP005024	AP005024 Streptomy
19	23	4.6	347496	1	BX842577	BX842577 Mycobacte
20	22	4.4	723	8	AK063543	AK063543 Oryza sat
21	22	4.4	1019	8	AK107232	AK107232 Oryza sat
22	22	4.4	137312	2	AC135598	AC135598 Oryza sat
23	22	4.4	349737	1	BX572597	BX572597 Rhodopseu
24	21	4.2	727	1	AF144053	AF144053 Micromono
25	21	4.2	39314	1	SGR300302	SGR300302 Streptomy
26	21	4.2	47981	6	AX112026	AX112026 Sequence
27	21	4.2	47981	6	AX112026	AX112026 Sequence
28	21	4.2	138203	1	AY310323	AY310323 Streptomy
29	20	4.0	3161	8	AK110400	AK110400 Oryza sat
30	20	4.0	3835	8	AK112081	AK112081 Oryza sat
31	20	4.0	3900	1	AF262754	AF262754 Amycolato
32	20	4.0	4116	8	AK121779	AK121779 Oryza sat
33	20	4.0	7080	1	AF305325	AF305325 Comamonas
34	20	4.0	7853	1	MAU66560	MAU66560 Mycobacteri
35	20	4.0	12097	1	AE005116	AE005116 Halobacte
36	20	4.0	17589	1	AE004862	AE004862 Pseudomon
37	20	4.0	20235	1	SERERYAB	SERERYAB
38	20	4.0	20235	6	AR049368	AR049368 Sequence
39	20	4.0	20235	6	AR095529	AR095529 Sequence
40	20	4.0	20444	1	SEERYABS	SEERYABS
41	20	4.0	31422	6	E38021	E38021 Avermectin
42	20	4.0	31422	6	BD097650	BD097650 A method
43	20	4.0	41156	1	MSGY222	MSGY222
44	20	4.0	43280	1	SFU78289	SFU78289 Streptomyce
45	20	4.0	48000	1	POL505006	POL505006 Sorangium

ALIGNMENTS

RESULT 1  
AMM223012  
LOCUS 53784 bp DNA linear BCT 09-FEB-1998  
DEFINITION Amycolatopsis mediterranei genes encoding rifamycin polyketide  
synthases, ORFs 1 to 5.  
ACCESSION AJ223012  
VERSION AJ223012.1 GI:2764760  
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.  
SOURCE Amycolatopsis mediterranei  
ORGANISM Amycolatopsis mediterranei  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.  
REFERENCE 1  
AUTHORS Schupp,T., Toupet,C., Engel,N. and Goff,S.  
TITLE Cloning and sequence analysis of the putative rifamycin polyketide

synthase gene cluster from Amycolatopsis mediterranei

Unpublished  
2 (bases 1 to 53784)  
Schupp, T.  
Direct Submission  
Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma  
AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,  
SWITZERLAND

Location/Qualifiers  
1..53784  
/organism="Amycolatopsis mediterranei"  
/mol\_type="Genomic DNA"  
/strain="LBG A3136"  
/db\_xref="taxon:33910"  
1336..15543  
/function="CoA ligase, acylcarrier protein,  
b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"  
/note="ORF1  
PKS modules 1-3"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAAL1035.1"  
/db\_xref="GI:2764761"  
/db\_xref="GOA:O54666"  
/db\_xref="SPTREMBL:O54666"  
/translation="METDILKPLHVLLENATRFACKPAPADHRTVYGDLEARRR  
LAGHLAGLGRHGRDVAICLGRNYSIVESFALLRAGAVGVLNPGSATALEHPLTDD  
SGATVYDAAQARLRLAPHLVBLVGTGDPGEGANSYDEHALSEPAEPAADLELDE  
PAMWFTSGTGRPGKVGSTQRCNLMSVASCYVFPGLSQDQRLWPLPLFHLSHIA  
CVLSATVGASVRIADGSSADDVMRLIEASSTFAGVPTTHLVRARQGRFSAPS  
LRLTAGGAVIAGLASEPTEFGVPLIDAYGSTETCGATTNPPDGRVGGSCGLAV  
PGVDVRVDPDGLDVPAGEGWSVGNVLMGYHNSPEATAAMRDGWFRTGLDAR  
RDDAGFTTCGRTELKELIRGGANIHGEVEAVRTVDGVADAVGVPHDTHLEVPVA  
YTPGPTGFPALAIKCRQLSAYKVPDRILEVAHRTPTASGKIRGLLTDPAQIR  
YAATHEEQSRHDESVAALRALRGLDERAQCELEDLVRTQAAADVGLQVPPDGRA  
FRDHTGLSLAIENRLRTEHTGLMTLPASAVFDHPTPAALARVRAELLTQVQVAP  
VVAADCEPTAIVGMACRLPGVSPEDIMWLVAERVDVSEFPDGRMDLSDLDPD  
REAGTSYVQGGFLHDAGEFDAGFGISPREAVAMPDQORLLLETLSWEALNAGVDP  
IALKIDTIDTGVGLMGQYSGSNAVPLEGFTVTGTVASSVSGRVSIVLLEGPAVTV  
DTACSSLVAMHLAAQALRGECMSMALGGVTWNATPGSFVSFRQALAPDGRCKAP  
AAAAGTSGEGVGVVLEISVARERGHRLAVIRGSVAVNQDGSNGLTAPNLSQQ  
RVTRRLAAAGLAPSDVVEAHGTTLPTLPTLVDKRLTEVMDWSAGAEILLTEARWPRNG  
IGHQAAGVAGVIMQVQALRHETTLPTLVDKRLTEVMDWSAGAEILLTEARWPRNG  
RPRAGVSFGVGTGNALHILEAPAEVPAPELPPVPLVVSARSTESLSGOAERLA  
SLEGGVSLTEVAGALYSRRVILDERAVVAGSREAVTGLRALNTAGSGTSGKVWV  
FPQGGTQWAGNRELLAESVFABRIACAAALAPWIDNSILVDVLRGEGDLGRVDVLQ  
PACFAMVGLAAWVESVGRVDPADVHSGHEIAAACVSGALSLEDAKVALRSQAIA  
AELSGRGMASVALGEDDVSVLVDGVEVAANGPSSVITAGDAHALDALTLEILSGEG  
IRVRVADVASHTRHVEDIRDTLAETLAGISQAQAPVPFVSTVISEWRDAGVLDGG  
WYVNRNVOYRFGAAATALLBOGHTVFVESAHPTVQPLSELTGDAIGTLRRDGGI  
RRLASMGELFVRGIDVDWTVMPAAGVDLP.TTYAFERHRYLBPAPASAGDPLIGT  
VSTPGSDRLTAQWRRQAPWADVLGVNFAALVEAARILGLDGLTPTVVGELVDAAP  
VLPFRSGREVQLIVFSGEQRRRP.IEVFSREADEPWRHAGHTLAPAAAVPPEAAA  
GDAVTVVAGLDADRYGIHPALLDAVAVTVGGDILLFVSVVTSLLASGATAVTVTP  
TATGLRTDPAGQATVSVRGTFPFAVEQTTDIALFRVDMPEIPLPTEATADLPYE  
ATSAEATLSAQMLADPAETRLAVVTDGTCPCGAAILWGLVRAQSGHEPVI.VIADLI  
DDPAVLFAVNASGEPQVRNGVAVSPRLTEVTPQDARPLDPGTVLITGCTGTIGA  
LTABHLVTAHGRHLVNSRGEAPELOEELTALGASVIAACDVADRAQLEAVLRAI  
PAEHLPTAVITHATGADGVVTELTPLDRLATVRPKVDARLLDELITREADLAPFLF  
SSAAGVLPNCPQAGVAAANAEILDALQRNSLDLP.FAVSIAMGYWATVSGMTEHLDV  
LRRNRICMSGLPADEGMALLDAAITGCTLVAAKFVDAALRATAKAGPVPVPLRGL  
APLPRRACETASITLERLAGAETEQAALLDLVRHAAEVLHSGAESVHSRTFKD  
AGFSLTAVELNRLLAATGTLSPAMT.FDYKPKPALADHLRAKILFGSAAANRPAEIGT  
AAAEPTIAVAMACFPFGVSHSPELWRLVADGADAVTEFFPADRGNDRLRIHEDDPH  
EGTTVTRAFGLDDAGDAAPFGISNEALAMDQQRLLLETSMELFERAAILDPTLL  
AQGDIVFAGVNSHDYNSMRHRAAGVEGFRLTGGSASVLSGRVAYHFVGEPAFVDT  
ACSSLVALHMAVQALQRCESMALAGVMWMTVEITFVSFRQRLAPDGRCKAFAD  
GADGTGSEGVGLLVERLSAQRGHVLAVVRGSAVNSDGSNGLTAPNPSQORV  
TRKALAAAGLSTSDVDAVEAGTGTTGLDPTIEAEALLATYQNRTEPLWLSVKSNGI  
HTQAAAGVAGVIMQVQALRHETTLPTLVDKRLTEVMDWSAGAEILLTEARWVSNHP  
RAGVSSFGICGTGNHNVLEVAAPITTPQEPFAEFLVPLVLSARTAGLQGAGRLA  
APLGRDTRVDPDAAYALATTRAQLDHRVAVLASDRAQLCADLAFSGSVVGTGTPVDG

CDS

CDS

KLAVLFTQCGSMAGMGRLEAETFPFRDAFEACEAVDTHLREPLREVDDEALL  
DQWYTOGALFAVETALFRLESWGVRPGLLAGHSIGELAAHVSGLVDLADAGELVA  
ARGRLMOALPAGGAMVAVQATEDEVAPILDTGTVCAAVNGPDSVVLSTGEAAVLAD  
ELAGRKTRRLAVSHAFHSPILMEPLDDFRVAERLTRYRAGSLVSTLTGELAAAL  
SPDWVGVRNARVFSDAVATLGAQSTFLELPGCALAAMALGTGLGPEQSCVATL  
RKNAGEVDVLTALAEHLVRGVGVDMVTVDLDEPATAVGTVLPTAYAFQDFWVDVET  
AAVSVTPPAEPVVDPRVQDVULLEVRSAANVLVHRDAGSDPLDRSKDHGDFDSAV  
KLNRILRDTFTGVELPSTLIFDYPNPAVLADHLHAEILGERPAPAPVTRDVSDEPIAI  
VGMSTRLPGGADSPLEELWKLVAEGRDVSGFVDRGMDLGLYHDPDHAGHSTYRS  
GFLHDAQFDAGLFGISPREALAMPQORLLLETSEALERAGVDPPLSGSDVGVFT  
GIVHVDYTRLEVEVDQGYTMTGTAGSVASGRVAVYVFGEGPAPVTVTACSSSIVA  
MHLAAQALQCECSMALAGATVMSADPALEFSRQGLSADGRKAYAGADGTGWA  
EGVGVVLERLSVARERGHRLAVIRGSVAVNQDGSNGLTAPNPSQORVIRGALASA  
GLAPSDVDVHGHTGTALDPIEQVQALITATGQEREQLNLGSLKSNLCHTQAAGV  
VGVIRKIMAMRHGVNMPATLHVDERTSQVDSAGAEVLTEAREPRTPRRAGVSSP  
VAGALVSGRATLNERAVIAGSDEAQLQALARGENAPGVVGTAGTGPKVWVFP  
GOGSOMNGRDLDDSSVPFAARIKECAALQWTDWLLDVLGRDADLLDRVDVQOP  
ASFAMVGLAAVWTSIGTVPDVLHSGQELIARAACVSGALSDDDAKVALRSQALAG  
ELAGRGMSVALSDEDAVUTPMANRVEVAVNSPSSVVIAGDAQALDEALEALAG  
DGVRRVADVAYASHTRHVEAETLAKTLAGIDAKVPAIPFTVILGTVITVQAVRDD  
GYWENLRQOVRFGPSVADLAGLHTFVEIISAHFVLQPLSEISDDAVTWSLRDD  
GGLRLLSAAELVYRGVADVMTVAAPAAQWVDLPTTAFDRHRHHEATAEAGVM  
DGEFWTAIBQSDVLSIAELLELPQORGALSTVVPVLAQWRDRRERSTAEKLYQVT  
WQPLREAGVPGGRWLAIVPAGTTDALLKELTGQCLDITVLEIIEASRAQAQORN  
VLAEDHLTVLSLLDGGPADAEETITASTLALVQALGTTTTSAPLWCLTSGAVNIGI  
QDAVTPQAQAVWGLRGAVALERLDRWGLVDLPAALDARTQAQLGLVINGAAGEQL  
AVRSGVRRRLVRKEPVTSEFCADTDRDAIQAQVADSEVTVAVVHAADIQTSVDDT  
TGTDGVDLTELAEAREKVTSEFCADTDRDAIQAQVADSEVTVAVVHAADIQTSVDDT  
GVADLDVPAKVTVTAWLDQPLDPAFVVFSSIAGIWWGGGQGPAGANAVIDA  
LWEWRARGKATSIANGLDQIGIOMDEAALARRRGGVIMPAFLAVTAMVQAVAG  
NEKAVADMWMAAFIPATTSVPSPFLPADLPEAKALLRAAQDDGEGDGTSSIASDL  
RAVPDEQRILLKLVRGHASTVILHSGAEGIGRQAFQVEGDFDSLAAVNLRSIHRNA  
TGLRLPATLIFDYPTPEALVGLVRELLREADDLGRDDEDLRSLVAFARFKKAG  
VLDLTGLADTGTEPGCTEATTEAPADDAELIDALDISIGLVQALGQTS"  
15550..30759  
/functions="b-ketoacyl-ACP-synthase, acyltransferase,  
dehydratase, ketoreductase, acylcarrier protein"  
/notes="ORF2  
PKS modules 4-6"  
/codon\_start=1  
/transl\_table=11  
/product="rifamycin polyketide synthase, type 1"  
/protein\_id="CAAL1036.1"  
/db\_xref="GI:2764762"  
/db\_xref="GOA:O52789"  
/db\_xref="SPTREMBL:O52789"  
/translation="MANOSWRKNMSADNEQIVDALRASLKENVLQENSALAAAAE  
PVALVSACVYAGGIRGPEDFWRVVSAGADVTFGPEDRCWDVGLXHPDPDNPETTY  
VREGAFLODQAQADGADGPGISPREALAMDQQLLEVSMTETLERAGIDPHSVRGSDI  
GVAGVQHDYADPLSGFEGFMSLERALGTAGGAVASGRVAYTIGLEGPAVTVDMGCS  
SLVAIHUAAQALRRGECMSMALAGSTVMTATPGGVPFARQALAFDGRCKSYAAAAAG  
LSAGLTPSDVTVEGHGTGTGLDPEVQALITATYQGRDPPQPLWGSYKSVVGTG  
SMAEGVGVLLRLLSVARERGHVLAVIRGSVAVNQDGSNGLTAPNPSQORVIRKA  
QAASGVAGVIMQVQALRHETTLPTLVDKRLTEVMDWSAGAEILLTEARWPRNGHPR  
GGISVAGSGTNAMHILEBAPDEDPVTEAPTVPLVVSAAATAAIIAAQAGRLAE  
GKVSADVAGTLVSMGRAMLSERAVVAGSHEEATGLRALARGESAPGLLSGCRSGVP  
DRVDVQOPFASFAMVGLAAWVESVGRVDPADVHSGHEIAAACVSGALSLEDAKVAL  
LRSQIAASLARGGMSVALSDEDAVUTPMANRVEVAVNSPSSVVIAGDAQALDEALE  
EALDALLDQGVRRIRVADVAYASHTRHVEAETLAKTLAGIDAKVPAIPFTVILGTVITVQAVRDD  
VEDAGVLDGGTVYRNRRLRRASAEELFVSGTVVDMGSDVFPFSPRRVSELPITYAFDHQHWLQ  
AVVTGTRDEGGLRRLLSAAELFVSGTVVDMGSDVFPFSPRRVSELPITYAFDHQHWLQ  
PGTVVLDLRAQDLGADHPLIGAVPLPQSDGLVFTSRLSLKSHFWLAGHAGTGVLLI  
MGTVDLALRAGDLGVLLEIVLEIAPILVIGERGVRVGVAVSGNETSGRADVDF  
SMREDEGWTHRTGLLGASTSRPSRPFPAAPPAEPIDVENFYDITLTERGVAYS  
QAFQGRVAVRERGRVAVLPPDHREDACKFGLHPALDLAALHTNAFANPDDRSV  
LPPFWANGLHVAQASLRVAVACGGPDALTFOQADETGGVITVMTDSLVSEVSAOL  
ETAAEERDSLVQVDMTEVPATETAETAEHAELEFAFGAAPIELTDDGDFLAVLAS  
AADARLIVVTRVAVREYTDPAQAVMGLVRAAQANEPRIILTDDGDFLAVLAS  
GEPOLAVRGNAPFRLARTEGPEPAFVSPGECTVLLTGTGSLGLVAKHLVLA  
GYRILVLSRSGVAAEDLVTELTQGAQTVSVVACVSDRQVAAALABHRTPTGIVHLA  
GLLDDGVICALNRELAVFAPKVDVAVQHDLDELDTGLDGLDAFVVPSSAAALMGSAQGG

NYAANAFDGLMAGRAAGLPCVSLANGLAEQADGLTANLTSATDOARMSRGVLPMT  
 PAELADIIDI GLAECALLVPIKLDIRLURQATAGGEVPHLLRGLGRASRVTTAA  
 ASGGGLVRLKAGRAEAEQAVLLGTVQAEAAVLFENAPELAQGTGRFSDLGFSLT  
 AVELNRLSAAKTGVLPAFLVPTVQALAEHLREELGETVAGAPATPVITVADAGE  
 PIAIVGMALCPGVMSPDILMRWAEGRDMSPPFGDRWDLGLFSDLPERPGETAY  
 IKQGFLEHAAFLPDGFFGCSPREALAQOQRLLEASWEALERAGIDPTKARGDAV  
 GYFSGVSHIDYLESLSNMPABIEGFTVTTATAGSVASGRVSYTFEGEPATVTDACSS  
 SILVAHLAAOALRQGCETWALAGGVAVMGSPIGVIGMSRORGMAEDGRVKAADGADG  
 TVLSGEVGTWILERSVAREGRHVLAVLRSAVNDQASNGLTAPNGSQOQVIRSA  
 LAGAQVSEVDVVEAHGTGTALGETIEAQLIATYKSGREPLMLGSLKSNIGHTQA  
 AAGVAVIKWQALRODTLPTLTHVQETPQKQDMSAGAVELLTEGRENHGRPRAG  
 VSGFSGTNAHLILEAPADDTAREADVPDAPVAVVVSARSSTGSLAQAGRLAFLDG  
 DVLETRVAGALLSTRAITLDRVAVVAGSAEAEAGLTALARGESAGLVTGTAGMPGK  
 TVVFFQGTQWAGMRELLLEASPFAERIEECAAALQPMWIDMLSLDLVLRGEGELDRV  
 DLQPAECFAVMVGLAAVWASGVVDPVAVLGHSGVEITAAVCPASVVIADAEALTEAV  
 QATAELSGRGMASTQLSHDEVAARLAPWAGRETAAVNCPASVVIADAEALTEAV  
 EVLGGRRVADVASHYTHVEDIQTDLAETLAGIDAQAPVVPFYSTVAGVETDAGVVD  
 GGYRRLNMQGFGPAVFAELIEQGHGVFEVSAHPVLVQPISELTDVAVTGTLRDD  
 AGADHFLGLVLPQSDGLVFTSRLSVRTHPLADGVPAALVELAVRAGDRAGCPV  
 LADLTVEKLLVLPESGRLVQVIVSGERTVEVYSQLEGADDMIRNATGHLSATPAHE  
 AFDTMPAPAGAOVDGLMRGDEIFAEVALPELDAGFIHFDLIDAVSQVPLAD  
 EQPAWESLVHHAAGSALKVRLVPGCALQADETGLVLTADSVAGRELSAGKTRAG  
 SLTRVDTWESVADSAPVANTEVVEAFGEPLTGLTGLVLEAVQTMVLTAAADRLVYV  
 TRGAVREVPDPAAGAVNGLVRAQAENPREFILDTDGETIPALGTDEPEIAVRGCKFF  
 VPRITRAEPGAAVFRPDGTVLISGAGALGGLVARRLVEHGVKVLASRRGRDADG  
 VADLVADLAADVSVACDSDRAQVALLDDEHPTAVVHTAGVTDAGVLETDRDLA  
 TVAPKVDVAVHLDLDRDLDAFVYSSVFMFGAGSGSYAAANFLDGLMANRR  
 AAGLPGLSLAWGLMDRQGTMAAGTDEATRAMSRGGLQINTQAEQMDLFDALSSAE  
 SLIVPAKLDLGRVADRAAGGVPHMLRGLVRAAGRAQAARASTVDNGLAGLAGLAPA  
 DQTLTDLVLRVAALVCHADASAVRVDTPAKGPDLSLTAVALRNRMTATGLKLP  
 ATLVDVNPQALAHRLDELGAQTPVTTAAKADLDEPIALVGMARLCPGVGAGP  
 EDLWLVNAGDAVSSFTDRGWDTSIDYDPDPAKPTGKTIKHGFLHEAGLFDAGFE  
 GISPREAAMDPQORLLLEASWEAMEDAGVPLSLKNDVGVFTGMFGGVYAPGDSV  
 VTPELEGAGTGGSSVASGRVSVYFGEPAVTTDSACSSSLVAMHLAAQSLRGEC  
 SMALAGATVMANPAGAFYEFRRORGLAVDRCKAFAAADGTGMAEGVGVVILREVS  
 ARERHRLILAVGSAVNDQNSGLTAPNGSQOQVIRRALVSLAGLAPSDVDVVEAH  
 GTGTTGTGDPFIEAQALLATYKDRSPFLMGLSKNIGHAQAAAGVAGVTKMVOALRHE  
 VLPTLHVDRPTPEVPLVAVVVSARSASLAGQAGLAGFVAGDASLAGAVAKALVTRAAAL  
 APAREPVPPTPEVPLVAVVVSARSASLAGQAGLAGFVAGDASLAGAVAKALVTRAAAL  
 TERAVMVVGSREAVTNLEALAGEDPAAVVTGRAGSPGKLVVWVFFQSGWSQWIGMRE

Query Match 100.0%; Score 501; DB 1; Length 53784;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-220; Indels 0; Gaps 0;  
 Matches 501; Conservative 0; Mismatches 0

QY 1 GCGGTCCCGACGCGCAGAGAACCGCATCTCTGTAAGCTGCTCGCGGCCACCGTTGCG 60  
 Db 15043 GCGGTCCCGACGCGCAGAGAACCGCATCTCTGTAAGCTGCTCGCGGCCACCGTTGCG 15102

QY 61 ACGGTCTCGGCGCACGCGCGCGAGGAGGATCGGCCCGCGCAGGCGTTCAGAGAGGTC 120  
 Db 15103 ACGGTCTCGGCGCACGCGCGCGAGGAGGATCGGCCCGCGCAGGCGTTCAGAGAGGTC 15162

QY 121 GCGTTCGACTCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCG 180  
 Db 15163 GCGTTCGACTCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCG 15222

QY 181 CCGCTGCGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCGCTACCTG 240  
 Db 15223 CCGCTGCGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCGCTACCTG 15282

QY 241 CCGTTCGAACTCTCTGCGGAGGCGCGACGACGGCTTGACCGGGGGGAGAGACGACTCGG 300  
 Db 15283 CCGTTCGAACTCTCTGCGGAGGCGCGACGACGGCTTGACCGGGGGGAGAGACGACTCGG 15342

QY 301 CAGATCTCTCGCGCGCGGCTGCGCGCTTCAAGAGGCGGCGGCTGCTGAGACAGCTG 360  
 Db 15343 CAGATCTCTCGCGCGCGGCTGCGCGCTTCAAGAGGCGGCGGCTGCTGAGACAGCTG 15402

QY 361 CTGGGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 Db 15403 CTGGGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15462

QY 421 CCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 Db 15463 CCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15522

QY 481 GCGCTTCGGGCGAGACGAGCTGA 501  
 Db 15523 GCGCTTCGGGCGAGACGAGCTGA 15543

QY 421 CCGGCGCGCGCGAGACGACGAGAACTGATCGACCTGACGCACTGACATCTCCGGTCTCGTGCACGA 480  
 Db 15463 CCGGCGCGCGCGAGACGACGAGAACTGATCGACCTGACGCACTGACATCTCCGGTCTCGTGCACGA 15522

QY 481 GCGCTTCGGGCGAGACGAGCTGA 501  
 Db 15523 GCGCTTCGGGCGAGACGAGCTGA 15543

RESULT 2  
 A69720 LOCUS 53789 bp DNA linear PAT 07-MAY-1999  
 DEFINITION Sequence 3 from Patent WO9807868.  
 A69720 ACCESSION  
 A69720.1 GI:4774328  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM Schupp, T., Toupet, C. and Engel, N.  
 REFERENCE 1 (bases 1 to 53789)  
 AUTHORS RIFAMYCIN BIOSYNTHESIS GENE CLUSTER  
 TITLE Patent: WO 9807868-A 3 26-FEB-1998;  
 JOURNAL CIBA GEIGY AG (CH)  
 FEATURES  
 source 1..53789  
 /organism="unidentified"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32644"

ORIGIN  
 Query Match 100.0%; Score 501; DB 6; Length 53789;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-220; Indels 0; Gaps 0;  
 Matches 501; Conservative 0; Mismatches 0

QY 1 GCGGTCCCGACGCGCAGAGAACCGCATCTCTGTAAGCTGCTCGCGGCCACCGTTGCG 60  
 Db 15043 GCGGTCCCGACGCGCAGAGAACCGCATCTCTGTAAGCTGCTCGCGGCCACCGTTGCG 15102

QY 61 ACGGTCTCGGCGCACGCGCGCGAGGAGGATCGGCCCGCGCAGGCGTTCAGAGAGGTC 120  
 Db 15103 ACGGTCTCGGCGCACGCGCGCGAGGAGGATCGGCCCGCGCAGGCGTTCAGAGAGGTC 15162

QY 121 GCGTTCGACTCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCG 180  
 Db 15163 GCGTTCGACTCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCG 15222

QY 181 CCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCGCTACCTG 240  
 Db 15223 CCGCTGCGCGCGCGGCTCAACCTCGGCTGAGCTGCTCGCGGCCACCGTTGCGCTACCTG 15282

QY 241 CCGTTCGAACTCTCTGCGGAGGCGCGACGACGGCTTGACCGGGGGGAGAGACGACTCGG 300  
 Db 15283 CCGTTCGAACTCTCTGCGGAGGCGCGACGACGGCTTGACCGGGGGGAGAGACGACTCGG 15342

QY 301 CAGATCTCTCGCGCGCGGCTGCGCGCTTCAAGAGGCGGCGGCTGCTGAGACAGCTG 360  
 Db 15343 CAGATCTCTCGCGCGCGGCTGCGCGCTTCAAGAGGCGGCGGCTGCTGAGACAGCTG 15402

QY 361 CTGGGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 Db 15403 CTGGGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15462

QY 421 CCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 Db 15463 CCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15522

QY 481 GCGCTTCGGGCGAGACGAGCTGA 501  
 Db 15523 GCGCTTCGGGCGAGACGAGCTGA 15543

RESULT 3  
AF040570 90445 bp DNA linear BCT 16-APR-2001  
LOCUS  
DEFINITION  
ACCESSION  
AF040570 AF040571  
AF040570.2 GI:12331604  
KEYWORDS  
SOURCE  
AMYCOTATOPSIS MEDITERRANEI  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Mycolatopsis.  
REFERENCE  
1 (bases 1 to 90445)  
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R., Yu, T.W.,  
Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X., Hutchinson, C.R. and  
Floss, H.G.  
Biosynthesis of the ansamycin antibiotic rifamycin: deductions from  
the molecular analysis of the rif biosynthetic gene cluster of  
Mycolatopsis mediterranei S699  
Chem. Biol. 5 (2), 69-79 (1998)  
AUTHORS  
98174059  
9512878  
2 (bases 1 to 90445)  
Kim, C.G., Yu, T.W., Fryhle, C.B., Handa, S. and Floss, H.G.  
3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the  
formation of the precursor of mC7N units in rifamycin and related  
antibiotics  
J. Biol. Chem. 273 (11), 6030-6040 (1998)  
98165773  
9497318  
3 (bases 1 to 90445)  
Yu, T.W., Muller, R., Muller, M., Zhang, X., Draeger, G., Kim, C.G.,  
Leistner, E. and Floss, H.G.  
Mutational analysis and reconstituted expression of the  
biosynthetic genes involved in the formation of  
3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin  
Biosynthesis in mycolatopsis mediterranei S699  
J. Biol. Chem. 276 (16), 12546-12555 (2001)  
21201076  
11278540  
4 (bases 1 to 90445)  
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,  
Hutchinson, C.R. and Floss, H.G.  
Direct Submission  
Submitted (30-DEC-1997) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
5 (bases 1 to 90445)  
August, P.R., Tang, L., Yoon, Y.J., Ning, S., Mueller, R.,  
Hutchinson, C.R. and Floss, H.G.  
Direct Submission  
Submitted (19-JAN-2001) Dept. of Chemistry, University of  
Washington, Box 351700, Seattle, WA 98195-1700, USA  
REMARK  
COMMENT  
On or before Jan 22, 2001 this sequence version replaced  
gi:2792340, gi:2832263.  
FEATURES  
source  
1. 90445  
/organism="Mycolatopsis mediterranei"  
/mol\_type="genomic DNA"  
/strain="S699"  
/db\_xref="taxon:33910"  
423..1451  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01706.1"  
/db\_xref="GI:2792310"  
/translation="MAPTGGVSGSRYKTLRQLEVALIGLIALRLHLEWLGAAGV  
RVYDGRSRLAEFGAAGSLEALDGAVIDICTPTASHEIATLTAAGVGV  
CEKPLASAEAEIVTAAEFAGVRLVAHDVRAAPVARLHLEVASRGLGAGRF  
SFSAYHPTWTHASASGGILITQLHGLADLHAWFGDVVRVHACVQGDITAPBG  
AVATGVTLTHASGIAISOVSRWTATPRPVVRVAFVSGVSGSVSDSEWFQVRVD  
GGAGNFAYGGPSVFTDTEMREFATAPAGPEPRIKAKDALAAVRIHAAASAWTGRAV  
ELPVKGA"

CDS

1448..2215  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01707.1"  
/db\_xref="GI:2792311"  
/translation="MKVAILSSSTPOAYAGALRGLPDVEVVAASWDAEPFVQQAABAG  
ARVLCYPPAAKETDLKAMILDADAADGDLTFAPACHGEAFVVRKGIADGGIGELTTL  
GSVATSDVGLGAAAPYLLDLADAVLGGEPAQOQVTAQTNIVLSRIGESAALVTVYR  
SGOVASFQRRHGSATGLPANTFIDQGSVOYDAGPOLLGGERPBLGGEDLEALMLKD  
ELGAGDGRGPDQAALFRIRIIQAAYESAHTGQPVDL"  
2353..2619  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAC01708.1"  
/db\_xref="GI:2792312"  
/translation="MQDRSLVGDPPKQVFAAMSVTQVKIALSGGPAELAPAHRSVDSG  
ALENTLKIRHAGAYEHFVHGGEFQAVDGDIDIAVFRWSHRTKIAE"  
2652..3842  
/note="ORF0; similar to Streptomyces Sp. Chop"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 monooxygenase"  
/protein\_id="AAC01709.1"  
/db\_xref="GI:2792313"  
/translation="MTDAISFEVPMWDRDKFPPAVFDSLRERPLAKMVYPDGHVGM  
IVSSYELVRELSLDFSHSCVCEVHPVTHQGOVIPTPLIPGMFTIMDPREHTRVRK  
LITGERTVRSRLIPRAEAVAAEQIEVMRAKADPADVMDPAKPIVLKMLGELVGLP  
YBERDYPVAVTLLHDAEDAPAEAAAYEVAGKPFDEVIERRQRQDDLLISLVIED  
LTQEELRNVTLLLFAGYTEGALATGVFALLHHTDQLAARAEPEKLDALPELLLR  
YLTQNYHTVTALEDVLEGELEIKKGDVTVTSLPANRDPKAFGCPAELDIDRTSG  
HVAFGFIHQCLQNLARTELRAGFTALLRAPELRLAVPADEVPLRLKSGVFSVKKL  
PVSM"

gene

3883..18090  
/gene="rifa"

CDS

3883..18090  
/gene="rifa"  
/notes="Rifa"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAC01710.1"  
/db\_xref="GI:2792314"  
/translation="MRTDLIKPLHVALLENATRFAGKPAFADHRTVTYGDLEARTRR  
LAGHLAGLVHRGDRVAICLGNRVSTVESYFALIRAGAGVPLNPGSAETAELEHPLTD  
SGATVVYDAAQAARLAPLPHVLLVTGDDVPBGHSDYDLALSEPAEPAADDLELDE  
PAWFTYSGTIGRPKGVSTORNCIMASVCYPPFGLSQDQDRVLPFLPLFHSLSHA  
CVLSATVVGASVADIGSSADDVRLIEAESSTFLAGVPTTYHHLVRAAKQKGSFAS  
LRIGAGGAVLGALRSEFEETFGVPLIDAYGSTTCGATMNPFPDQARVEGSGCLAV  
PGDVVRVDPDTGLDVPAGEGEVWVGNNVLGYHNSPEATAAAMDGWFRTGDAR  
RDDAGYPTICGRKELIRGGANTHPGEVAVLRTVDGVADAAVGVPHDTLGEVPVA  
YAATEHQSRHADESVAALRLSLDERAQCELLDLVQTAADVLQVDPDGRA  
FRDGLTSLAIVELRNKLTHTGLNLPASVAFDPTPALAARVABELLGITQVASEP  
VVAADGPEPTAIVGMACRLFGVSPEDLRLVAERVDVASEFPFGGWDLSLIDPD  
RERAGTYVGQGFGLHDAGDFAGFGISPEAVAMDPOQRLLLETSWALEAGVTV  
IALKGTDTYVSGGLMGQGYGSAVAPLEGFTVTGASVASGRSVSVLIEGPAVTV  
DTACSSLVAMHLAAALROGCSMALAGVTVMATPGSFVSQRORAPADGKCAF  
AAADGTGNSGEGYVTVVLRISVARERGHILAVLRGSAVNOQGSALITAPNGLSQ  
RVIRALAAAGLAPSDVDVVEAGHTGTTLGDTIFEAQALLATYGERKQPIWLGLSKN  
ILLEGDVLTEVAGALVSRRAVLDRAVAVVAGSREEAVTGLRALNTAGSTPGKVVWV  
RPRRAGVSFGVGTNAHLILEEAPAEPEVAPPELPVPLVWSAGAIELTETARAWPRNG  
IGHAAGAGVAGVTKMWQALRHETLPTLTHVDKTELDVSWAGAIELTETARAWPRNG  
SLLEGDVLTEVAGALVSRRAVLDRAVAVVAGSREEAVTGLRALNTAGSTPGKVVWV  
PFQGTQWAGMGRLELAESEVFAERIAECAALAPWIDMSIIVDLRCEGLGRVDVLQ  
PACFAMVWGLAAWVESVGRPDADVHSGHSEIIAACVSGALSLEDAKVALRSQAIA  
AEISGRGMSVALGEDDVVSRVLVDGVEAAVNGPSSVITAGDHALDALTLEILSGG  
IRVRYADVASHTRHVEDIRDTLAETLAGISQAQAPVFFYSTVTSWDRDAGVLDGG  
RRLLASMGELFVRGIDVDTAMVPAAGVDLPTVAFERHFWLEPAPAGAGDPLLTGT  
VYSTPGSDRLTAQAQSRPAQPVADGLVNAALVEAAIRLGDLAGFPVVGELVVDAP  
VVLPRGSRVQLIVGEPGQRPRPIVFESEADEPWRHAHGTLLAFAAASPEPAA  
GDATDVTVAGLRDADRYGIHFALLDAAVRTVVGDDLLPSVMTVSGLSASGATVTVTP







```

/codon_start=1
/transl_table=11
/product="glucose 1-dehydrogenase"
/protein_id="AA054105.1"
/db_xref="GI:21449373"
/transl_table="MPRLAEQVWVWVGVASSGIGRTALRFAAGARVCAARNTAL
DLGEVRCAGGAVATADIAEAAVAVADVAVRFGVDTWVNAAGIGVYGVED
TPAGEFVRVNVYLGHVGHAKALPALRAGGVLLGVASVLGLRSAPLQAPYAAK
Query Match      6.2%; Score 31; DB 1; Length 82746;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 GGCACCGGCTGGCGTCCCGCGCGCTG 198
      |||||
Db 17833 GGCACCGGCTGGCGTCCCGCGCGCTG 17803

RESULT 5
AF016585 41097 bp DNA linear BCT 07-DEC-1997
LOCUS Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nida)
DEFINITION gene, partial cds; polyketide synthase modules 1 through 7 (nida)
genes, complete cds; and N-methyltransferase homolog gene, partial
cds
ACCESSION AF016585
VERSION AF016585.1 GI:2558836
SOURCE Streptomyces caelestis
ORGANISM Streptomyces caelestis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces; Streptomycetaceae; Streptomyces.
1 (bases 1 to 41097)
Kakavas,S.J., Katz,J.L. and Stassi,D.
IDENTIFICATION and characterization of the niddamycin polyketide
synthase genes from Streptomyces caelestis
J. Bacteriol. 179 (23), 7515-7522 (1997)
98053867
9393718
2 (bases 1 to 41097)
Kakavas,S. and Stassi,D.
DIRECT SUBMISSION
SUBMITTED (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park
Road, Abbott Park, IL 60064, USA
FEATURES
    source
        1..41097
            /organism="Streptomyces caelestis"
            /mol_type="genomic DNA"
            /strain="NRRL-2821"
            /db_xref="taxon:36816"
            complement(<1..396)
            /gene="nidi"
            complement(<1..396)
            /gene="nidi"
            /note="ORF6; similar to tyli"
            /codon_start=1
            /transl_table=11
            /product="cytochrome P-450 hydroxylase homolog"
            /protein_id="AAC46023.1"
            /db_xref="GI:2558837"
            /transl_table="MVDVSGPMELSKDANAKELDFSHNRTHHPVFWDEGQAWOV
FRYDYITVSNHPFFSDFTFVAPTPELEMLIGPTIGALDPAHGPMKLYSQAF
TPRMAGQEQIRVIAEELLDVRGQKTTA"
        627..13649
            /gene="nida1"
            /db_xref="taxon:36816"
            /note="ORF7; similar to tyli"
            /codon_start=1
            /transl_table=11
            /product="polyketide synthase modules 1 and 2"
            /protein_id="AAC46024.1"
            /db_xref="GI:2558838"
    gene
        627..13649
    cds

```

```

/translation="MAGHGDAATAQAQAEKSEDSGDAIAVIGMCRFPAGPABTAEFW
QLLSGDAVVTAAADRRRTTIDADFDAAFFGMSPREAAATDPQORLVIELGWSFL
EDAGIVPSLRGEASVFGAMNDYVATLHRAAGPTDYATLGLHNLINRSLVFL
GLRGLSVDTQSSSLVALAVESLRGGTSGTLAGAGVNLVIEGSAAMERVCAL
SPDCHHTDARANGVRGEGGAI VVLEKLADALADGRVYCVWRGVAIVGNDGGGCL
TPYDGRGQAVLRAACDQAGVPADVPVELHGTGTPAGDPVAEALAGVIGTGRPAN
EFLVAGVONTGHLGEGAGIAGFKAAALCHERLALPASINFEETNPALPLERLRUK
CTAAALOPKTGGGCLLAGVAGMGTCHVLEBETPGCGTQADACILFSASP
MILLASRQALRAQAARLREHLEDSGADPLTAVSLATTRFEHRAAVPCGDPDL
SSALAALAGOTPRGVIGSTDADGCLALLFTGQAQHPGMGOELVTTDHPFAALDE
VCEIQRCTQNLREVMFTPDQDLDDRTETPTQALFALQALYRLTARGTQAHVL
GHSVGTAAHIAHAGVLDLPDAARLITARAHVMOQLPHGGAMLSVQAHEHMDQLAHT
GVEIAVNGPFTCHVLSGPTEALREHREQNRVHTLWKSVAHSHAFSLMDMLGAFR
DTLNLNYQPTPLIISLNTGQIADPNHLCTPDYMDHARTVRFADAVTAHGTTC
TYLEIGPHTLTLHHTLNDPTTITLHREPEPTLTAQIAAGVTRDGDIAVLC
GRASPRVELPTVAFORRTWAPGLTPNAPADPAAPORAMAVGDPVLELPTFR
ETIASVIGDGDPEVALDRPFTSQGLDSMTAVELAGLGTAGVAGLDPVLELPTFR
AVADHLAKTLGESAADQOVNRTGEAEAKGDPIAVIGTCIGRFPFGVATPDDLWE
LVASGTDATSTPTDRGDLGIDPDPTPGKSYVRHGGFTLHDAQFDEFFGISPR
EATAMPQORLLILETSWEALERAGVPESLRGRTGTVFTTAPETGPRLEHGTGYE
GFLTCTTASVAGRIAYALGTRPALTVDTACSSSLVALHVAQLRREGCECLAG
CTTWSGRCMVFEFSORGLADPRCKAFSADAGTAWARGVGLVLESLDAERLHG
RVLAVRGTAIVNQDGSNGLTASGPAQVIRDAISDAGLSADDDIDAVEAGHTGAL
GDPIEAGALLATYHPRKQTPVWLSKSNIGHTQAAAGIAGIIMKVOALRHDTPEK
LHADHPSSKVDMDAGELQLLTDAPWPDPRRAGISAFVSGVNAHVLEPAP
EEAPATQAPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPDQTAHALATTTHF
KHAIVTGRTRTELHTLTDALIOQTAAHPRLTLFTGQAQHPGMGOELVATDPH
PAAALBEVCELRQCTQNLREVMFTPDQDLDDRTETPTQALFALQALYRLTARG
TQAHVLHGSVGTAAHIAHAGVLDLPDAARLITARAHVMOQLPHGGAMLSVQAHEHMD
DQIAHTHGVEIAAVNGPFTCHVLSGPTEALREHREQNRVHTLWKSVAHSHAFSLMD
PMLGAFRDTLNLNYQPTPLIISLNTGQIADPNHLCTPDYMDHARTVRFADAVT
ADQRTTTLTLEIGPHTLTLHHTLNDPTTITLHREPEPTLTAQIAAGVTRDGDIAVLC
TLHHTSPOTHLDLPTYPEQDRYMEVVRVAVSQGQADRLRYRVLVWMAPEHPT
VPSARWLLGSGSDPADAGLIVGVEQLVGHGAVOELDIPAPAGRLLPSCSLPYR
GPSTGEFCGRSLRSPRVAEVALLOALQADAGADAPLMIATRCARAAQAQEAESVGOA
QMLGQVAGLELADRWGVLVDLPADAPALRLKARLVSNADPNVALRAGSVFVR
RVVPAPGRVPRNDWAPSGVTITGCTGALGSQVARRALAGAPHLHLLRSGVGAEE
LVDELTAIGAETVAACDAARDALVSLATIPEHRLTAVLHAGVLDGVDLGLTF
ERIDAVLAKATAARHLDELTDLDLDFVLFSIVGVWGNGOATVAANAALDALA
HGRARGORATSIANGPWAGSGMAGGAKSLARDGIDVLEPEALDLQAVRSDST
TLIVADVWETELSSASRRFPPLFGRIFAARAVASGADGSPVRLVWLSQLSAA
ERQLRLDLVTEASVIRHGSATVDAETAFRAGGDFDSITLVELNRDLVSLGRLS
NTLLFDHTPAVAEVLGEQLLGTPEPAAASHGADAGARTVVASPVVAVMGACPYG
GVALPDDBELVASGTDAISTFTPDGMDLGLYDPPDPTPGKSYVRHGGFTLHDAQF
DAEFEGISREATAMPQORLLILETSWEALERAGVPESLRGRTGTVFTTAPETGPRLEHGTGYE
QSRBAADYCGHLLTGTLSVMSGRVATLGLQGPALTVDPTACSSSLVALHVAQSLR
RGECDLAAGSTVMSFTVFRSROGLAPDCRCKPFPADTADGTAWARGVGLVLE
RLSDAERLHGRVLAIVRGTAIVNQDGSNGLTASGPAQVIRDAISDAGLSADDDID
VEAHGTGTPILGDTIEAEALATYHPRKQTPVWLSKSNIGHTQAAAGIAGIIMKVO
AMHGTLPRLHVDAPTQVDWEAGSLQLLAARFPADPRRAGISAFVSGVGTNA
HVLLEPPAPEAPARTQAPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPDQIA
HALATCTHPRKRAVITGRTELEHTLHLDALIOQTAAHPRLTLFTGQAQHPGM
QOELYTTDHPFAALDEHICELQRCGTQNLREVMFTPDQDLDDRTETPTQALFALQ
ALYRLTARGTQAHVLHGSVGTAAHIAHAGVLDLPDAARLITARAHVMOQLPHGGAM
LSVQAHEHMDQLAHTGVEIAAVNGPFTCHVLSGPTEALREHREQNRVHTLWKS
SHAFSLMDMLGAFRDTLNLNYQPTPLIISLNTGQIADPNHLCTPDYMDHARTV
TVRFADAVTAHGTTC TYLEIGPHTLTLHHTLNDPTTITLHREPEPTLTAQIAAGVTRDGDIAVLC
LATHTHTGHTTTPHPSHPAQRVSLPAYPPFORRAYWMSAAHIGRSDAERATRLGLA
QDPHPTGATSI VSGSGLTLTRVSLAAHFWLADHTVAGAVLFGTAPADILLHVAE
ETGSGTVEHTLHVPLVPAQAVQLOVLVEAPDERGRRATVAAARPEETGAETHIT
AAGVLAGEAPAPADWADWAGAWPPFGAEAVDDELAYEAFVGYDGFATGLTG
AWRGEFVADPRLVPEAVAGSEERRFGVHPALFDAAHPHVRGVPWQESGPDADGG
QSTLLPFSQOGLTLVGAEOVRVRLAPSDGGAFAFVRAADAPHPVLDALVLRVP
SSTGALPAAPMHIDMQPVDSTSLVTRRVAVVAGDADETAELVLAFLPDHHTTAHA
DLAARAPADAVPTGPVALLPGSPPPAPATSPADPVRATLHAGLSLIVQALIE
DERLTGTTTAAVTVTRGAVATQTDVDPDLAAGLWGLVNSQSEYVPGRLVLDVDSPE
TGALRALADEPQLARAGVMLAPALTFAAAGSPAPADAATAHGMDGTVLITGC
TGALRRVAVHLARRHGVRLMVSROPNAPESAALRELTGLVHTSFVSCDIADP
ADVRKAVAGVPSDHLGVVHTAGVLDGALARTPERIDTVLAKADAVRHLHEATL
GHPLRAFVLFSAAAGLGRPCQVAAANAVLDARLRRRAEGIPAVSLAWGLMDERE
GMAGGLDATALRRRLRGLPMPADQLELDQALITRDGFPALLVPLULLSTALURR
AGENGASVPLLRSLVSPRHRHTTATSVRDESGPGSPGSAALPAKSRMTVLLD
TVSEQVAGVLGHAASAEIDPRPFHEIGFDSLTIELNRLGLRVLGLPLSLAPDHP

```

TPRKVAEWINAEIPANPALGTSAEDPVLITGIDOMARAIALMDAEDGRRADVRORLIVG  
LITLALDRPASGADSTRAATTVADRILDEATDDEIFAFLDEQLEEQDPDELGAFLBERL  
DEQP"  
13693..19212  
/gene="nida2"  
13693..19212  
/gene="nida2"  
/note="orf2"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase module 3"  
/protein\_id="AAC46025.1"  
/db\_xref="GI:2558839"  
/translation="MTAENDKIRSYLKRATAEHLTKERLAELESARHEPVPVAVGMAC  
RYPGVATPDDILWEIIVASCTDAISFPTDRGMDLGLYDPSDTPGKSVRGGFLHD  
AAQDAFFGIGSPRENTAMDPOQRLLLETISWEALEHAGIAPNLTTRTGTGTGYMD  
DYGSCFFSLPEVGVVNGSVASGRVAYTLGQGPALITVDACSSSLVALHIAV  
QSLRGECDLALAGGTVMTPTVFYFQRGLAPDGRCKGFADTADTAWAAGVGV  
LIVRERDLAERLGRVLAIVRGTAIVNQDGSNGLTAPSGPAQQOQVIRDAAGLSAD  
DIDAVEAGTGPLGDP1BAGALLATYGHPRQTPVWLGSLKSNIGHTQAAAGIAG11  
KMOAMHGTLPRTLHVDPSPKVDWKAGSLLOLAQRPWADPDRPRAAALDSWLG  
GTNAHAIIEBPSTLIEAPVDETPVPAHTPPVIAWPLSAHTTALRAQAARLDSWLG  
GTNAPODIAHALATHTHFKRAVITGRITKLTHTLIDALOGTAHPRUTLLHT  
GQAQHPGMEGLYTTDPHPAALDEVCLEQRCSTONREVNFTPDQDLDRTEYT  
QOPALFALQALYRTLTARGTQAHVLGHSGVEITAAHIAAGVLDLPDAARLITARAHL  
GOLPHGKMLSVQAAEHLDDOLAHTHGVEIAAVNGPHTCVLSQRTALBETAQHLEQ  
NVRHTWLVKSHAFSALMDPMLGAFTDLTNTLVNQPPTPLISNLGTADPNHLCTP  
DWDIARHTVRPADVQTAHQGCTTYLEIGHPTLTLLHHTLNDPTIPLIHEH  
PEPETLTALATLHTTGTTTLHTTSPQSHLADLTPFPQDRDWMVAPRAAGVOLA  
AAGVATVHPLLAATHEVETDPTVTWMLSALEPMLADHLVMDRGPAGTVALEIV  
LETRIGLIGLCEGLVEAPLSWGDAPAVVTVVGLSPGTRVLSYSRPDEAVW  
THASGLRAGLSEEMSGSVFAELSGVWPPGGAQEVDSIGVELFAAGVGGVGAAPRG  
LEAVWERDEIPEAVELPDODHADAAGYGLHPALDAALHPIALLDPLGDDGHLGPP  
VTVLVRPNGHGTILVRVTRAQDPADGGQDRPGGAVSLAAADRQGVVPVARSILV  
RVTSATLPAAFPVPLHRLDWPETPGFTSTGAGSDPSVLDLTDLTADVLHTRVRL  
EAVQARLPFGDGLVLTTRGAQEPGGADAPAGAAWGLVRSQAQSHPGFGLID  
VDTESRSALPAVMAAGVEDVAPVAVQRPSPVLPLVLTADRTEAEPGFGSAGL  
LGPDTGLVLTGGTGAALAVTARHVARHARHLLVSRGTDPAQAEALASLGAALGA  
DVEFRACDVTDAVORLLAG1IAEHPILSCVHTAGVLDGVTSAQTPERVAAHPGX  
ADGALHDELTRDRERSVPLVLFSSVSGVLGSGHAGYAAANAFALALARRADGHP  
ALSLGHWMTGTGMAGLRODARTARRSGIAPLDAETAMDLPDRALARPEPALVPR  
LDHALREARATLEVLRLDLAGVPFSAHPAPGSEGTATSTGTGSLAERLARLTVA  
ERLPLELIVTEVAAVGLHGDGPGVGAERSFKDAGFSLTAVDLNRNLNARTGLRIP  
ATLVEDPTPLALAEILLAGLVEPADGPEIDAAAFDLRLARGLAALDRAARVGAPE  
RLSLTLITLSDADPAGTGAQGEADGALTETVTDRLRSASDDELFDLYDSER"  
19269..30233  
/gene="nida3"  
19269..30233  
/gene="nida3"  
/note="orf3"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase modules 4 and 5"  
/protein\_id="AAC46026.1"  
/db\_xref="GI:2558840"  
/translation="MSATNEKEKREYLRRAMGDILHSARLERLAELESARHEPVPVAVGMA  
CRYPGVATPDDILWEIIVASCTDAISTPTDRGMDLGLYDPSDTPGKSVRGGFLH  
DAQDAFFGIGSPRENTAMDPOQRLLLETISWEALEHAGIAPNLTTRTGTGTGYMD  
HDYGSQGVGAADAGGQUGTGVASGRVAYTLGQGPALITVDACSSSLVALHIAV  
VQLSRGECDLALAGGTVMTPTVFYFQRGLAPDGRCKGFADTADTAWAAGVGV  
MLVLRERDLAERLGRVLAIVRGTAIVNQDGSNGLTAPSGPAQQOQVIRDAALAHLP  
DIDAVEAGTGPLGDP1BAGALLATYGHPRQTPVWLGSLKSNIGHTQAAAGIAG1  
IKMVAORHDTLPRTLHADHPSSKVDWEAGLQLOLAQRPWADPDRPRAAALDSWLG  
SGTNAHAIIEBPSTLIEAPVDETPVPAHTPPVIAWPLSAHTTALRAQAARLDRHRTSDAL  
SLADSGIGHGLVAGRAALPHRAVLILGDRAPLDAALASGEVSPDVVTGSRADVRK  
VAFVPCQAGQWAGMGAELLDSSVFAAELARCEALPEFPVWMSLTDVLRGAPGL  
DRVDVQVPTFAVVVAAAMWGLVEPAAVVGSGEIIAAHVAGVLSLEDAARVVA  
LRSQLIARELAGRSMASVAAADAVESRLAGAGGGRDVEIAAVNGPETTVVCGA  
PGVDSLGLVLCQGVTVR1DVVYASHSRHVGIRDELAAVLAGLPRAGRVPFYST  
VEAPLDGTALDAGYVTRNLRQRVFSALRAMLADGVDAFVCESPHPVLITVPVQTL  
EDAGAGAVGSLRDRDGGRLRFLTSAEAQVAGVPVDMALCPRAGWDLPTVAFOR  
ERYVWPAEPGPGAGASAAATGPAATAARSGSQEDASAYAGDRLAYHVAWGGLIP  
GADAWRPGPRLI1VPAGEREAENVADSVEQAISFGGTVRVTRVTDPLSTGDRGLAILIA

gene

CDS

gene

CDS

PVYAGEEPLTGAVALLLGLCTDGHPEHPGVPAGVGTGLTALVQALADLGVAVPLWTVTRG  
Query Match 5.6%; Score 28; DB 1; Length 41097;  
Best Local Similarity 100.0%; Pred. No. 0.027; 0; Indels 0; Gaps 0;  
Matches 28; Conservative 0; Mismatches 0;  
QY 119 TCGGCTTCGACTCGCTGGCGCGGTCAA 146  
|||||  
DB 34720 TCGGCTTCGACTCGCTGGCGCGGTCAA 34747  
|||||  
RESULT 6  
AF521897 6567 bp DNA linear BCT 10-MAR-2003  
LOCUS Streptomyces hygroscopicus ansamycin biosynthesis gene cluster,  
DEFINITION partial sequence.  
ACCESSION AF521897  
VERSION AF521897.1 GI:28894473  
KEYWORDS  
SOURCE Streptomyces hygroscopicus  
ORGANISM Streptomyces hygroscopicus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 6567)  
AUTHORS Wang, Y. and Gao, Q.  
TITLE Cloning of ansamycin biosynthetic gene cluster, partial sequence  
from Streptomyces hygroscopicus 17997  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 6567)  
AUTHORS Wang, Y. and Gao, Q.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUN-2002) Pathway Engineering Dept., Institute of  
Medicinal Biotechnology, 1# Beijing Tian Tan Xi Li, Beijing,  
Beijing 100050, China  
FEATURES  
source  
1..6567  
/organism="Streptomyces hygroscopicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1912"  
gene  
1..2414  
/gene="shnC"  
CDS  
1..2414  
/gene="shnC"  
/codon\_start=3  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AA061320.1"  
/db\_xref="GI:28894474"  
/translation="EEDNVRLRVVGVSTLRRDEGPRRLLTSMALFVRGVVDWTK  
FLMAGHVDLPTAFDRRHYWIQDAATADDSGASNDADADFWSAVEQTDADSLAGL  
LAPDSAGLRDLRTVVPALADWKRGRSRSSAERLYAVTWRLDREVSVPAGRWLIA  
VLPPGCAETVTSRVAELIAELGAQGLDVPPETVPSAFTRTGLTARLSDIARVEYOP  
AGVLSLLADGEQDVIDTVARTLALVQALGDAGVNGPLMCLTRGAVNTGLQDTAGEEG  
DAAIWGLGRAALDEHPRGGLDLPATADAHTAQHVLGALNGTAGDQLAVRRPGLYS  
RLVRVQSTPADGWRPHGTVLVTGGABALGHASLWLRSGARRLLVITTAQAPA  
DAVTELOQLAAGVETTVVSCADADRETLARIETPREQPLTAVTHAADAPWTSVA  
ADTGHDLTEVFAGKVDTAVLDELFTGIDAAPLDAFVVFSSITAGVGGCGQVSGAA  
GAVLDALVRRGRGLAATSIAWGLDGLGLGMDAAALAAALRRRGLVLPMAHVAVDITAF  
EQAEAREKAVTVADMNEAP1PAPTSARVSLPADLPAAALRSQDPDAENGDLTS  
SLVDSLRVPQAEONRLLKVCQAATVLGHSSGESIGPLQFQEVFFSGLGAVNLR  
NSLHVATGLPATLVLPYPTPAVVGFTVSELTTSTDDLEGREDLRLVLAQVPLR  
RLREAAGLLDILLGLSDVSDGVPPEAXAPAPAPAAADAARDRDRGRPRKARSQ  
QPLT"  
2458..5433  
/gene="shnD"  
2458..5433  
/gene="shnD"  
/codon\_start=1  
/transl\_table=11  
/product="type I PKS"  
/protein\_id="AA061221.1"  
/db\_xref="GI:28894475"  
/translation="MTAPDEQIVDALRSLKENRLQEQNRLESSEASBP1AIVSMAC



```
EGPLGAMWGLWEOTGTMTANLSQADQARMSRGGLPIGVAGRMRLFDLTALRMPALT
AVPIKIDLRGMDAAGGVSILALRALVQTGRQQAQAGENGSLTARLASLTPOEQE
ELLIDLVLTOATVLTGYTEISNVEAHKAEAGFSLTSVELNRLRATGLKLAATS
VFDYTPPLALRHLHDLGPKAQSPGTGTHPLLAELSKLEARLVETPVDDDSIRAQVTRL
QVLLTTWSTANGTPTGDEALDFAISDDDLFLDITDTFSN"
9111..13136
/contig_start=1
/transl_table=1
/product="ShnB"
/protein_id="AAP45193.1"
/db_xref="GI:31211265"
/translation="MLELBAEKSPVAIVGMACRFPDGVASPDQLWRLVSEGREGLSG
FPEGKWDVEGFDADPXRKGYTDTRGGFLHAGLFDAGFGISPREALAMDPOORL
LLEVSMEALREADIPDGLKSDGVFSGVITQGYMRGETPELGGXLTGMXAXSVA
SGRVSVFGEFSGVSDTACSSIVAMHLAQAALRGECSPMALAGGATVAMPXFTFV
EFSRQXXSSDGRGXAYSADQGTWAGCAGVVVLELSEARNKHQVLAIVRSANV
QOGANGLTAPNGPSQORVIRQALASAGLSAADVDVEHGGTGVLDPIEAQAALLAT
YGOEREPRLMIGSVDLQQAALPIFYFTVTGEWREAGVLDGGYRNLRQVRFPGAIA
WGAGAVELLTEARDPDTGRPRRAVSSFSMGTNAHLILEHSPQAESVAEPVSDV
VPLVVSARSLIIVGQAQLASVGSASDVALPEVAAGLVSRRAVLSERAVVGASRD
QALSGALARGESHFVVTGSLGAGVGRKVLVFPQGGSPWGMGRLELCSAVFA
ARVGCALVELERWMSLVDLARGDCSVEFEREDVROPASFAVMWGLAAWESGVFA
ADAVVHSGGEVAAACVSALSIEDARVIVAVRSQIISGVLGRGMAVGLSERAAV
ARLQMDGREGVIGAVSPSVISGTEALQGVLDALAAGVSAARMEVDFAHSGRQV
EAEELIREAFADIRAQPLIFPYFTVTGEWREAGVLDGGYRNLRQVRFPGAIA
DLIADGHTVFVSSASPALWLINEIADANTEVLSGSLRFGDGLHRLFTSMADLF
VGRIVDMSGVLAAQADARRVDLPTVAFEHVWLPMTGTTDAASLISGADHPLIGA
LVIEAPLVPAQXKRVQXAVGCPENGXRSVAVYSARDDBGKGTGXDXDWTTHAGTL
TAAVPAQDGFDTWPPFGAERVSFADVFIEEMAGRGYVGYFAGFQVWRRGEEV
FEVALPDPQHEARSGFLHALLQGLVLRPADAGDMRVFPFAMNGLRLHAAGA
SELVRTPSPGSDAYSQAADGAGGVLSELSIVARAVDEQLDRMATDGRDALFEV
DMSLEPAPASVESIPPSALVAEDVTDLADAAPVAVPAVLEAVGCGDEHDLALITV
RVLEVQWAFAAAGLAESRLVVVTRGAVFVGEGEVADPAGAAVWGLVRAQAENPDR
IVLLDLAADVDMGSLVPSKCG"
13847..14533
/codon_start=1
/transl_table=1
/product="ShnB"
/protein_id="AAP45194.1"
/db_xref="GI:31211266"
/translation="MPLPSPAKESKMTAEKLLSKLVTAELHDLRQVARTDEBPV
AIVGMACRLPGGVSGQDILWRLVSEGREGVSGPEGRGWDVEGLFDADPGRACKS
RGFLHAGLFDAGFGISPREALAMDPOHXLACGCTGPGCGLPRWRCKSSWTG
GVLVMPORAGVPRCCVLCIPVGGRRRPGRTAVELPRGSPVSHRRSRKHCCSTW
RWSASATSR"
14963..15505
/codon_start=1
/transl_table=1
/product="ShnB"
/protein_id="AAP45195.1"
/db_xref="GI:31211267"
/translation="MPLPSPAKESKMTAEKLLSKLVTAELHDLRQVARTDEBPV
AIVGMACRLPGGVSGQDILWRLVSEGREGVSGPEGRGWDVEGLFDADPGRACKS
YTD RGFLHAGLFDAGFGISPREALAMDPOQWYRWCTHPASRRSHSDRASSRSVSL
VSPRTGLPFPVAVSTPMAPR"
15358..16812
/contig_start=1
/transl_table=1
/product="ShnA"
/db_xref="GI:24637553"
/translation="MGVLADVHASGPOVALRSGAVLVRARVADTDRASTCRRLLDP
DGTALITGTCGALVAVRHLVVEHKIRSLVLSRGPDPAGADLDLAELTALGARVR
IVACDIADRAAGELIASVPRDAPLTAHVHTAGVGGVVTALPERLDARVPRADA
IVLDELTRHLDVAAPLFFSSARGTGNGPGGNLAASNAVLDALAVRRRTAGAPTSV
ANGVQDTGTISGLGVADQSRMARWGLAASQAGLELFDALAXDADVLVAAXXNFA
GLRAQASBPVHLLRGLVGRAGRAQAQASRREGLAGOLATPPVQREQLILLDLVR
EVAALVGYTPKVRPDPAQDVGFTSVLAVELNRLAGLIGRLPQASIAFDPTPRR
WMRHLAELCEPGSPADREDEIRALATTPLSRPFELGLMEQLHLVHAPSGSAA
APDTAEFKQDAGPLIEMDIDNIVKRAEMEKARP"
16813..22131
```

```
/codon_start=1
/transl_table=11
/product="ShnB"
/protein_id="AAN63813.1"
/db_xref="GI:24637556"
/translation="MDARQNTIEIIVASSKSVKVEALRASLMENRLESEVQIRDSLT
EPIAIVGMACRFPFGVSPSEELWELIADGRSAVEAFPTNRGMDLENLDLPDRPGTT
YVREGAFLHAGFEDAGFGISQSETVMYMDPOORLMLETSWAEFRAGIDPAAMRGKN
VGVFAGMAQGYGTAFHSIIPDELEGYVMTGLASVLSGRVSYTTFGEGPAAVTIOTAC
SSSLVHMAQSLSGSESSIALVGGTNMATPTAFVLTARAGGLAKGRCKKAFASA
DGTNWAEGVILLERLSDAVNRGDEVILGVTRATVAVNQDASNGLAANGPSQQVRIR
QALAAAGLSPADVDIVEAHGTGTGLDPIEAQALLTYGRNRPAGLPLWLGSKNLG
HAGAAGVAGVITKMMWMMRHGVLPTLTHVDETPEDVMSAGAVELLTEAHEVPYGR
RRAAGSGCAGCTNAHVLLEQATEPSTGNLPDEKARVLGDSVYVPLVVGARGKAGLQ
AHLRGSFTQRODITDVLIGOSLRSRGNLPDEKARVLGDSVYVPLVVGARGKAGLQ
GVVTFARSTVGRVTLVFPQGTQWAGMAELLEASVPFAARMTCACVLDPLTGMSL
LDVVEQREGARSLDVLQVSWALMYSLAALWEACGVVPVAVVGSIGEIAACYA
GALSIPDARLMVHRSLAEALVGRGNASLTADVKAIVSLIETWPGSEIETAAVNGPA
SVVVTGELPSLEELLARCEADGIRARRINGAAHSSQIDVLHDSFVEALASVAGA
SRVPLYSTVTRGLHDTTTFDVEHFRNMQTQVDFAIRSLVGDGSGVFIYSAHPVL
TSSVDVLADLEAGPAVVTGTRDDGGPRFLASLHLYTHGVGDWEAVLGRGGEQ
PVDLPTAFORORYLETAESRGDAPGLGLEVANHPILGAVTEIPGSDGLVLTSLSL
RTHPLADHAGAGVILLPGAAVELAVRAADEVGYLGVSELVIERPLVLPESGGVQR
VMVGPDSEGHRTVQVHRSREAGSKSWTRHSGRLVPEDCQAEFLQWPPPGATA
VDPDAFAHYDLHAEGYHYGPAQOQRAAWTRGEVFAEVSLSPEAGKADYGLHRA
LDDAAMTSLFRPDLSDSPKLALPVWMDVRLHADGASTLRHLTLPADTIRLHLA
DTSQTPVASVSLVLRPVVPELLRVSGAAQDMFRVAMEPLSVRSVDDELKARVTT
AEDVRAAATAPRVLLDVAGDRTDPAARDLSGRVLEAVCAQMLAEPAFOPTLLAL
THSGAARVGDGDPVPLAVATAAGLLRAAQSENVGRILLVTDGTGTEASARLLPDVLAAG
```

Query Match 5.2%; Score 26; DB 1; length 22131;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 TTCCAGGAGTGGCTTCGACTCGCT 134  
Db 2008 TTCCAGGAGTGGCTTCGACTCGCT 2033

RESULT 8  
AF521085 132544 bp DNA linear BCT 23-MAY-2003  
LOCUS Streptomyces nanchangensis NS3226 polyether ionophore nanchangmycin  
DEFINITION biosynthetic gene cluster, complete sequence.  
ACCESSION AF521085  
VERSION AF521085.1 GI:31044135  
KEYWORDS Streptomyces nanchangensis  
SOURCE Streptomyces nanchangensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
REFERENCE 1 (bases 1 to 132544)  
AUTHORS Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
TITLE A Complete Gene Cluster from Streptomyces nanchangensis NS3226  
JOURNAL Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin  
REFERENCE 2 (bases 1 to 132544)  
AUTHORS Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-2002) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

FEATURES  
1. .132544  
/organism="Streptomyces nanchangensis"  
/mol\_type="genomic DNA"  
/strain="NS3226"  
/db\_xref="taxon:204925"  
gene complement (682..2052)  
/gene="nanti"  
CDS complement (682..2052)  
/gene="nanti"  
/note="putative integral membrane transport protein"





WVLSARTADALRAQAQGLAELWVTAGAPRSPASPTSPASPADVGVWGLIATRSADRHRAV  
VSGTRDELLSGLRAVADGLAFAAASAGAAPGVVMPFGQSGWGMVGLDSSPVF  
ARMAACEALGFEVDWLSITAVLRGAPGAPERSVDVLOPCLWAVVSLAAWESYGV  
TTPAVVGHSGQGEIAAACVAGGLSLRDCARVVALRSGALRALAGHGTMSALUSGAE  
REFLADGAAARVAVFNGPYSTVVSGETDQVAAVAAACEAAGHARTIDVDYASHG  
PQDLRLADRTITDLADLSPGADAVFYSAVTGARQTELDADYFNTLRQVFRASA  
IDALLAAGVRFVTEYSGPHVLIIPALRCEFEAEAAATVTLRRQGGPDQVARALGD  
GFVAGLAVDWSWFVTEGDEAGDEGRHRTVELPTVFFORRYWLPADGPDGREGRTAG  
VETPRAGHALLASAVELADGGLVLSGLPGDAWGAHTVAGVQVLPVGAVDWALLA  
ADAGASLEBELLRLAPLSGLSEPSAGVLAQVAGADDESRRRLRISRRPAD  
AGAGEWTHAVGSLAPGPPAPAUTGTATVMPVPAQVAGADDESGRRRLRISRRPAD  
PALRGVVALWRGDADVALPEEAGGEGGEGDGTAGFLGHPLVLLDADALPALLI  
ABPDGTGCGAGCEARLMLPFAMSGVRLWATGARARVRLSLDGGGGDVADERELRI  
EVSPTDPAVSLVRPRTVROVREASGAAGLFDLMTVPAPQEPGSAEDDAG  
CVAVLGAEPTEPCVDCRTYDLPALLAALDAGAPLPSVWVRPPAADPGAPADAA  
LSNVRGVAALRAWAEPLTISRLAVTVRGVAGAGSEGPVDLAAAANGCARGVQ  
AHPDRIVLVDDDDMGADTIDIGAAAGLAAAGLGFQVLRGDTLLAPLRASAA  
TEGGVDFNGTIVTDSGSLAGSVAEHLVRAEGVRLHLLVRFEGADGAYDTRDQD  
AQVHMTVPDRTAALERVVAQVDPAPHLTVGVHAGLADITSGAARGWAAVAGV  
RALHQATVAAALPSRVFTLSDAVADWGPAPAPERAAGAFCAAVTDVRRRAGHLGLOVA  
FGPWAADDGAGDSGRWTVGLGDRGLALLRAACRADRPRLVADIDTRALTAHPA  
HELPAALRTLGASASAGRAPRVRVRAAAPGRTTDMASRLVGLGPAERRRRLVLELV

Query Match 5.2%; Score 26; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGCTGCGGCTGCCCGGAC 194  
|||||  
Db 53661 GCCACCGGCTGCGGCTGCCCGGAC 53686  
|||||

RESULT 9  
AB070949  
LOCUS  
DEFINITION  
ACCESSION  
AB070949  
VERSION  
AB070949.1  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
SUBMITTED (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of  
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1  
Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail:ikedamc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,  
Fax: +81-3-3444-6197)  
FEATURES  
Source  
Location/Qualifiers  
1. 78210  
/organism="Streptomyces avermitilis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"  
/note="Actinobacteridae"  
complement(1..1254)  
codon\_start=1  
/transl\_table=1  
/product="putative lipoprotein"  
/protein\_id="BAB69301.1"

AB070949 78210 bp DNA linear BCT 11-OCT-2001  
Streptomyces avermitilis polyene macrolide biosynthetic gene cluster.  
AB070949  
AB070949.1 GI:15824136  
Streptomyces avermitilis  
Streptomyces avermitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1  
Omura, S., Ikeda, H., Ichikawa, J., Hanamoto, A., Takahashi, C.,  
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,  
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.  
Genome sequence of an industrial microorganism Streptomyces  
avermitilis: deducing the ability of producing secondary  
metabolites  
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)  
21477403  
PUBMED  
Ikeda, H.  
Direct Submission  
Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of  
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1  
Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail:ikedamc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,  
Fax: +81-3-3444-6197)  
FEATURES  
Source  
Location/Qualifiers  
1. 78210  
/organism="Streptomyces avermitilis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"  
/note="Actinobacteridae"  
complement(1..1254)  
codon\_start=1  
/transl\_table=1  
/product="putative lipoprotein"  
/protein\_id="BAB69301.1"

/db\_xref="GI:15824137"  
/translation="MGEALITLCITRSTRLAATAGVLGGVLAIVTALGGTSNAATSGNGA  
GNHPTSOARINQAGAEASDARIETTPQGTQVSGTNGHPSVTVTSGKLTVMTAVTA  
TGAETPGTSLTDTGTSKNGRLERATKYQIAEAGDKVGRSTTGSATITTVGANDPFI  
GHLSPEDESTGVAMPVTVNFKALSDKAQVSEIRVSSSSQORVUGVHNDHVRP  
PENYRPGSTVTVTLNRHGIQKTVFKIGRSQISLTDAKTKMTVRDGTIKTIDPIS  
SGSPHPTVNGQWISSEKFRQTHMGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI  
HGNVYVTSFGCEANTSHCGVLDVRGDSHGRQPAAFWFDHSHISVGDVVIVKNSBEKT  
MLAPNGLSDNWMSWVAGSPAT"  
complement(1256..2290)  
/codon\_start=1  
/transl\_table=1  
/product="putative transposase"  
/protein\_id="BAB69302.1"  
/db\_xref="GI:15824138"  
/translation="MPVEFLSHDOVSRVGREASAPTPGELOQFRLDAKALRSASRKR  
AFANRLGVAQVQCMVRMLGVFTEDLSVPEVVRFAAQQLGVDGEFAAYGARQNR  
IYHAMEIRDAYAIRRSFAAQAQVRFLATRWASWEGPRALFDRAVGNLVDNRLLPG  
ITTLRLVAEVAQENALYRTLDEAVPEKQSMRDLKVPGRVRSLELRRLRSPM  
GVSGADRLAFERIRMEAVMTAGPGSTDIKELRVSVQSVQRVRSGLAGDRPGRG  
AVPQVSVLSLTLFAALDEERAKGAPSWTITRYLHHDPPIFQGVYAKWSQCCLHSF  
GRSAQGYRDQ"  
2335..21081  
/gene="pceA1"  
2335..21081  
/gene="pceA1"  
/codon\_start=1  
/transl\_table=1  
/product="modular polyketide synthase"  
/protein\_id="BAB69303.1"  
/db\_xref="GI:15824139"  
/translation="MVSEELNATGNRDCSADSADGIAVIGLACRLPGAADPAFAWR  
LISGADAITDVPDRWDGAVADASAPRTGIRRGGLDSDIGDPAFGPLSPKEAA  
AMDPOQLVLELAWALEADAHVRAATLRTRTGTVFGVGMDDVATLHRSGLTASPH  
TVTGLHRSIIANRVSPFLGNGSPSLVDSGQSSLSVHLACESLKKSGESTIALAGV  
NLNVIPESALGAAGFGGLSPDRCFTFDARANGYVRGEGGIVVLKPLARAVADGTVG  
YCVIRGSVNNDDGGDLTVPLQSGQEVRLRAYERAGVNPADVPLHARVADGTPVD  
NFTGNPAIPLDLRLNRVOTHESEMAEDGRLLAGVSFGMGNTCHIVLAGASL  
ACEGATAGGGAAPTFWPLSAKTASGAQAQAAALHAETAPLGLALPDVGSLATGTA  
FEHRAVVGVRDDFLARLSTGIDDAITGRAGPRGELAFPLDSGQSQRAGMGR  
ELASVPAFASALDEVCAHLDPLPLPRREVFPAPVTGEAELDQTLTQTSLSFAVE  
VALFRLESNDITFDVIMGHSIGELAAHVAGVLSLADACALVAAGRMLQALPAGGA  
MIAQAQAEDEVRAIRGERTDRVISAALNGPDSVVSGDEDLATEIDAFAALRKTSR  
LRVSHAPSPHMEPLAEFRNVAETHTAPRIPISVNSVTRGRLSESPSEPTAAY  
VRHREAVPADGVLAEDQGVRTYLELPGGVLASMARAGDSLSFVPAALTRRE  
AQALLAAVSLHVHVEPDMDALFGGGRGARVVALPTVAFQRVWLDGDTTSPVAT  
TGWAPVAPQDAIETDSSAPVGLRRLAGLAESEDEAVLDMVRAHIAAHLGHAER  
QVETWTFKDLGDFSLSVLRNQLSEATGLRPLSPGLLFDHPTPAALHARLAAQSLLG  
GDAAAAPATVADPHEPIAIVAMSCRLPGVSSPEDLWRLLAGSGDGSIGFPTDRWD  
VETLYDPEPTGKSGTRHGGFLHEAAEFDAFGFISPREAAIDPCQRLVLETAWA  
FERAGIDPATLRGSRAGVFIGTAQDYGRLRHPADGVEGILLTGTTSVASGRVAYS  
FLIEGPATVDTACSSSLVALHVALQSLRSGECTMALAGVTVMSFGMEVRSRQSG  
LSADGRKAPAGADGTWAGVGMULLVERKLSDRRNGHVLAIVRGSANVQDASNG  
LTPAPGSOORVVRQALAAAGAPADVAVEAHGTGLTGLDPIEAEALIAAYGGDRG  
DRPLMGLSKNTGHTQAAAGVAGVIMVMNHGVLPTLTHVDEPTPHVDWSAGVVR  
LTPAEVWRPARDLPRRAVSSFGISGTHAVIVEQATSEPARGTAVAGVPWVWSA  
KSTPAVLEQAQGLLSFLDAGVGAGVREDVALSLATSRALERAAVVGGLDAELR  
QGLHALASGAGAPVCGVGVGFLFSQGSQRIQMGRELYAAYVFAAAYDEVCA  
FLDAPVDVDEELNQITGTQTPALFAVEVALFRLESNGRPLFDVAGSVEGEIAAHV  
GWSLEDAKLVLSARMLQALPDGAMVAQVATEDEVRHLFDVGVGIAVAGSRRSVV  
VSGAEDAVTAGVEIFAGRGKTSRLKSHAFSPMLDMLPEFAAVVRVTFSESRIP  
VNSNLGRLAEVVTPYVHREAVRFADGIETLGLBELVTEFVEVPGVGLAQAG  
CPDGTDI VTVPLRADRPEPYALVTAI GRLHTHGI SPDMQVLF PGARRV DLTPTAFQ  
RERYWLDAPRAATAGADAMDAEFNDSVESDRASIGALLGLEPTELDVVAKLSAR  
KKRREQSTDDGWRIRITWQPLDDMRAAPSGTWIYVSEETAETAI RAGLBELGVAL  
VPEFVAGDETDRLALAFADAGQEQADGILFAAQGEAGANALQRLVLLVQALGDAGI  
EAPLWCTISGAVSTGTADLTDPVSAQLWGLGRIVALEQQRWGLDLPFAELDSRAL  
ERLGLAQADEQLAVRAGVYGRVRLVRAPLSAAADAPWSPRGTVLTGCTGALGG  
QVAREWLAGAHEHLVLTSRGPDAPGAAELKAELELGAQVTVVACDADRDALAEFL  
GRYIVNAVHTAGVLDGLVRSITPERLNHVRPKVDAALHLHELTRDQDLDAFVL  
RRMTGVWNGGGQAYGAANAFLDALAEHRRRAQGVPAALAVANGSWAGMDGAGDHL  
RRSGVRAIALAPVPAISLHGLTGETSMTVADVDFWERFVPAFVGTGTRPCPLLGVPEAR



EALEBARAAQSPVPSVULVQRLGATPGCEQRIILLDVLVREQAATVVLGHTGKAFAE  
 DRTFRETGDSLFAVELRHINTATGLRFTTLVDFHPTPTALRHGLGELLGRHGLE  
 TAAQAATAAAADPEIAIVGMSCLPGGVSPDLWELLSSGSDAMSGFPTDGLWDID  
 ALVDPDPDRPKYTSREGGFIQCADRFAGLFGISPREALAMPQORILLLETAWFAPE  
 RAGIALASVAPSTGVPTGNGODYANGLENREEGYALTGKAASVSGRLSYTFG  
 LEGPAVTDVTFACSSSLVALHAAQAALSGECSMALVGGVTVTTPDLVFBSQRLGS  
 PDGRKAFAPGDTGWGEGVGLLVERLSDARRNGHOVLALVRGSAVNATDQASNGIT  
 APNGPSQQRVIRQALASAGITPADVAEAGHTGRLGDIIEAQAALLATYQGHSDR  
 PVLWGSKNSIGHTQAAGVAGVVKWLMRHHGVLPTQLHVEPTEHRAHVSAGVRL  
 TEVPOGTTGPRRAVSSFGIGTNAHTIIEBAPAAAAPAREHREPVPMVLSAK  
 SEVALRAERLLAFATDVSVDAGFSSATTSALHRAAVIGSDPAELRAGLEALA  
 AGPAAVNVAGRAHSDAGVLFSGQSGRLGMDYAAFPVFAAIVDEVCALLDAT  
 VVDVDETHTGTSQALFAVAVFALFELLESWGIRDYVAGHSVGEIAAAHVAGVLSL  
 EDAKIVSAAALAKMPLAGVVMVAQAAEDMPLHIDRAGTAAINGPRSVISGAE  
 DAVTAAEVAAQAKRTSLKVKSHAFSPMLDPMLEFAEVVRGLTFSEPMVIVSGL  
 TGRLTSTYTPYVVRHREAVRPAQVLTLELHGVTFVFIIPGQVLSALVQGLDDG  
 TDVTIIPALRADEPQALTAQAQHVSGAEVDMHAFEPGARREVPYTAQORERW  
 LDAPASAGDMRAVQGEAGHPLIGAANPLADSVGHLLTGRLSKTHPWLHVAVGVA  
 LLPGTAPELAIATAADAVGCDLLEELTLETPLVPERGGVALQVHVGAADGSGRRPLT  
 VHSQIDDDGDAYAHEDDSERAVRHASGLVLTATGPGVDGSLTANPPVGAEPVDPGY  
 DRLAGMALDYGPFQGLRAARSGDDPFAEVELPGAERTDAGAYGLHPALFDALHTV  
 WLGAVPEACTGNGLLPFAAGVRLAAAGASDLRVKVSRACTGTSTVSLVLAADTGP  
 VAEVDSLTLPVPADQLRAGSGSDAVRGLKMTVPVSLPAPDCMRLETVEDLEALRAT  
 DDPADTEADAAAADAADVPCPTGTADLATRVRETHAVLEVLVQWMLAEERFORLVL  
 VTRGDLAQSAGVLRSAQTPNDRIVLVETSDAHAGSGVDADLADARVLPGLIA  
 SEFPQAAVHADDEVVRPLVRVAVPDAEPGLGSGTVILLTASGSLGGLGFARHLVA  
 EHGVRSLIVSRREGGDAPGAELTADLTQAQADVTWAAACDVADRAAVRALLAGQOSL  
 SATIHTAGVLDDGIIGSLTPELIDAVFRPKVDAALNLDIAGELSGDLSAFVLSVA  
 CITLTGTQQAANYAANTFLDALAEHRAAGORATSLANGLWAQDSBSMTSGSLDHTLT  
 RIQRMGLAIIPSSDGLRDLFDALADRAVAPRLDGLGQGTVPVIRLGRVTRT  
 ARRTAQTPAGLQRLAGLAERBEQVLDLRTVEVAVLGHAGSGAGDADFSDI  
 GFDLTAVELNRNLNLAGLRELPATLIFDYNELTLAHFLTGEAAGSEATVAPFTLS  
 TGTGTADEPTAIWMAKRYPGVQSPEDIMQVLFSGDRDAGVSGPEGRGMVDEOLXDPN  
 PDQWGTSTYREGGLDADAEDAEFGISPREALAMPQORILLLETSWEAERAGIDP  
 ATVRSRTGVPAGVMWHDYGRVHTSPGPEGLVNGSAGSIASGVSYTFGLEBRV  
 VTUACSSSLVALHAAQAALSGECSMALVGGVTVMGSPVVFESQRLSGLDGRCK  
 AFGADGTGWAEGVGMMLVERLSDARRNGHKVLAVVGPANVDQASNGSLTAPNGFS  
 LKSNIGHSQAAGGVGIVKMMAMQGVLTPLHVEPTPHVDMSAGAVOLLTEAVEM  
 PASDGPRAAVSSFGVGTNAHVIVEQAATSEVTARGTAAGVPPVPLVLSAKSDAALRE  
 QAGRLISFLDAEGVGGVGPEDVAVSLATSRVLECRANVVEEGEELFQGLGELVIS  
 GAAPVAGVGRVGFIFSGQSGRIQGMRELIAYPFAAAAYDEVCALLDADLVDAE  
 TLHOTSOTPALFAVEVALFRLLESWGIRDYVAGHSVGEIAAAHVAGVLSDEADLI  
 VSAALMAQLPAGAGAMVAIQATEAEVLPHLTGHVGLAINGPRSVLSGSDAVALI  
 AVEYPAQGRKTSRLKSHAFSPMLDPMLEFEVVRGLTFSEPMVIVSGLNTRLAE  
 VYTPYVVRHREAVRPAQVLTLELHGVTFVFIIPGQVLSALVQGLDDGHTVTV  
 PALRADRPEYALVTAVQOLHTHGISPEQALFGARVLDLPTVFAQORERYNLDAQGA  
 SGVRAUGAAGDHLGALATADSDGAVTGLLSLGTHPWLAHNVGILVLLPQTA  
 FVELAIRAGETCGMIEDTLBELPLVPERGGVTIQVAVGPEDESGRRQLSVHSRVD  
 DRSWRHATVLSAEPASAKLGAWPAGAEAVDITAFYEGMAAGLDDGPVFGQLR  
 SWRMDGDVPAEVELPDSEAAEAFAHPALLDALHAAAGLSDSDSPALPFAW  
 SGVVQOAGAAVVRVLSRTATSGVTLAVADGAGDPVASVESLSLRAASAEOLPGAVG  
 GDQLFGLERTFTLPSATDALMETVADFDALRESEGTDPVVVPPVPPCAPSGRETA  
 DRVHSTISVIALVQWMLAEERAGRLAVTRPGDLAAHVAWGLVRSQAQSNPRIVLV  
 EASETEAVRLVLAAGSGPFAVRDGEVFPVRLVKATRSATGTATGTPDGVGGVCL  
 LTGASGALGVLAHLVAEHPQVLSLLLSRRGAEPGAVGELEAEALWGAEEVRAACD  
 VADREAVSGLMVLGEGQSLVAVHTAGVLGIVASITPERMREVPKVDVNLNHE  
 CTDRMDGLAVPSSVAGSVAGSQAAYAAANGFLDFAFSARRRREGPLATSLAWGYE  
 QSGMVTGLVBAARBARSGVPLPPEGLRFLPDALASDEAVLAPVRLDTGNARAG  
 EAPPVLRVPAARBARBARSGVPLPPEGLRFLPDALASDEAVLAPVRLDTGNARAG  
 GHASDRVTRPFAQDGLGFSLTAVELNRNLNKNKATGLRPLATVIFDHTPALLARHVF  
 MAAGAEPALPDVLSLLADRVFQELVTLTAAEARDRIILSLQGLLAIAGEFGKAPA  
 QEGFGTGGSDLESATDDEVDLLKKEFGIS"

gene

CDS

/gene="pteA2"

21273. .31847

/gene="pteA2"

/codon start=1

/transl table=11

/product="modular polyketide synthase"

/protein\_id="BAB69304.1"

/db\_xref="GI:15824140"

/translation="MSCRYFGVDSPEDLMELVAGGRDVAVSFPAADRGGWLDLALYDPPD

PERSGTSYAREGAFILRDARFAGLFGISPREALAMPQORILLLETSWELFERAGIDA  
 ASLAGSTGVFAGVNMHEFLAALQNSPQDELGYLGTGTSVAGSRVAYTFTGLEGPV  
 TVDTACSSSLVALHAAQAALSGECSMALVGGVTVTTPDLVFBSQRLGS  
 PFAAGADGTGMEGAGLLVERLSDARRNGHOVLALVRGSAVNATDQASNGIT  
 QORVIRQALAAQLTAAQVDAVEAGHTGELGDIIEAQAALLATYQGHSDR  
 IKSNIHTQAAGVAGIIVKMMARHGEPLRTHLVADPSVHVHWNTAGAVRLITERFW  
 SASDQRRAGISFGISGNTAHTIIEBAPAAAAPAREHREPVPMVLSAK  
 APFAAGPVPWISGTEEARAQBELLRSRTDAEVPPLDFAFLSATRTTALBHRVL  
 QGRTPALAEGLAVAGSGPARVARGLSTGTRGVGLFSGQSGRLGMDYAAFPVFAA  
 AFAAADEVCAOLDADVVDSEBELHTGAQPALFAVEVALFRLLESWGIRADYVAGH  
 SUGELAAHVAGVLSLDDAKLYSARALMQALPAGAMVAVQATDEVELPLHTDAVG  
 IAINPGRSPVVSAGEDATVIAABAFRQQRKSLKVSFAHSPMLDPIERAEV  
 RSUTHRPQIPVNSLTRLAEPTFYVVRHREAVRPAQVLTLELHGVTFVFIIPGQVLS  
 PGVLTAAGCLDDDDVTVPALRADRPERVALVAAVSELHAGHSVGEIAAAHVAGVLS  
 VDLTFAQVRYWLDASPAAPGGRVRAAGLSDHPLGAAVSLAGGDERLLTGRLSL

Query Match 4.88; Score 24; DB 1; Length 78210;

Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 CTGCCCCGCGACGCTGATCTTCGAC 207

Db 60409 CTGCCCCGCGACGCTGATCTTCGAC 60432

RESULT 10

AP005022/c

LOCUS

Streptomyces avermitilis genomic DNA, complete genome, section

2/30.

ACCESSION

AP005022 BA000030

VERSION

AP005022.1 GI:29603894

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

and J. Ishikawa.

- \*1 Kitasato Institute for Life Sciences, Kitasato University
- \*2 National Institute of Infectious Diseases
- \*3 The Kitasato Institute
- \*4 National Institute of Technology and Evaluation
- \*5 School of Science, Kitasato University
- \*6 Institute of Medical Science, University of Tokyo
- \*7 RIKEN, Genomic Sciences Center

Following url is also available.

<http://avermitilis.ls.kitasato-u.ac.jp>.

#### FEATURES

```

source
1..300425
  /organism="Streptomyces avermitilis MA-4680"
  /mol_type="genomic DNA"
  /strain="MA-4680"
  /db_xref="taxon:227882"
  /note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinus"
  /complement(382..645)
  /note="SAV258"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67967.1"
  /db_xref="GI:29603895"
  /translations="MNFALAPEVSVTDTEHGTVLDDERTGRYLLNGTGALRALRL
DGTSPQGVASLQCRFPDAGTAQVEQVDNLLHALRTAKAAS"
  /complement(642..2552)
  /note="SAV259"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67968.1"
  /db_xref="GI:29603896"
  /translations="MFAVRLPHLLLTIVESTGMFEVLVDPHPAGAAALALMPTAGPPP
TVIEHSGRSPVLVGNWEDVMVTAVGHRVALLGRTTADADVLTRRLRDVRSVRLD
DGLAAAGSFHIVASVNGEVRVQGLTSLAACQIYYATGVTVLAADRPQALADRIAG
IDEDVALALLSPFPGLPVSRLWRCVSPPTGCLRTGRDGTGTROWMTAPDP
RPLGDMITQALABAVATCTAGQVVVSADLSGLDSTSLCFIADADRDHDAHVTVHY
QALGRSDRYWARQAEALPGARHLVVGQAPDWYAPLAHAVHREGLPTVRSRA
TTEHLAAAGSRRLQIGGDELFPSPMALHVLRRHPALPLHLRLKSMRRR
PLTTARTLAASPSYQWLAAACSNFPADRDHWGADWEPPLKPLPMATPHAAATVRR
QLLAALATQPLATFPVHEMILRLTRIGNATRNRSIHGHTGVTFHAPFTDDRVLE
AILAVLEDRVALGQFKPLGTALRSIVPAGLLSRTTKGDASAEIYAGLRRRGELLK
LMDDSHLVRMGFYADVLRVTLRGLHADTRPLMPLDSTLACELWLRLTSLPADTARATA
VPVPSPTVKEI"
  /note="SAV260"
  /codon_start=1
  /transl_table=11
  /product="putative transposase"
  /protein_id="BAC67969.1"
  /db_xref="GI:29603897"
  /translations="MGRAGVAFITDEQRNNGYTFNQVDDGQIAGVFLDRDARRA
MTCGRARSQIYAGVLTGRTFGLFDLPEDAPAEVAYVAEQIHPGSLVAGYGAER
TRWDHQIATKDAYGYRDLKGDANKLRDGD"
  /note="SAV261"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67970.1"
  /db_xref="GI:29603898"
  /translations="MLGQGYVPLDLMIEGESORMSERDALGLRALMWVIAIEVPV
YLVLVSVLTGQQLPRGTREEDGKPRATD"
  /note="SAV262"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67971.1"

```

```

  /db_xref="GI:29603899"
  /translations="MQADAMAVAMELOQAAANRLIWGRGRTFTLTLIAFAGG
ARARITAGTDVQSGLVGQOAEELSRERVASKQTVAAVREPLAQAAATAAALMFCSS
APAVVAATQLLNALADVENTARLEALEAFGRVAVNATPPRLSWARRRVSAG"
  /complement(4811..6445)
  /note="IS6 family; completely same as SAV7545 and SAV75480
SAV263"
  /codon_start=1
  /transl_table=11
  /product="putative IS6 family transposase"
  /protein_id="BAC67972.1"
  /db_xref="GI:29603900"
  /translations="MILAECELLMTFPHLEGLVVEVSPGEGVLHIVARTVESVPV
PCDCATPSPRRHSGYQRLADGAVGRQVSIETVTRLLFCDDQRCARVTFAQVQDL
TVYGRRTQLRCLLSAIALAGAGERIARLPVVSRTTLLIALVMPALPDYATFP
RVLGVDEFATRGKHGYTVLDCETHAPIDLLPDRESATFAAMLTHDPGVEIICRDG
GADGARTCAPDAQVADLWHLNLAETVKSLSKSHSCLEPALSPEVILHP
PISHAGRLAERARRHAAVHDLQLQGLTIRAVARLELSRNTVRRYARVITWELATP
RWNLPSITLDPYKSYLHQWHEGTSGLAKHAELERGFTGSYSVVVDYLRFRRTSG
DRPPARPFGVRKVTGMITNPDNMNDQDKLKAILARCPELEAATGHVRSFAAWA
IRSASRLPEWIATARADEHGLRFPADGLIADLDIVILGLSTWSSGCGVEGRVTDIKL
LKRQAGRAGLPLLRKRVLLVAADRRQRHVNTQAH"
  /complement(6592..8802)
  /note="SAV264"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67973.1"
  /db_xref="GI:29603901"
  /translations="MTPATFTPPSSTPPSQCHGTTTPEDPACGRGIRVSGHTACIA
HUSDADRDAYLAGLSPGNVDHRTGTSFQDLLDALLDARDPDTPGPRLGVARFRSAT
FORDVFRSATFOKNADFGSATFQGNAGFLIATFEGDAGFSARFEGDAGFSARFEG
DAFVSATFEGDASFVSATFEGDASFVSATFQSAASLWGAAQCDVKFESATFQGDVAV
FGGADFPGDATFGLAAPERNAFLARFGLIASFLVTHGADWAFSAIFRGDASFVS
ATFQSDTLEAATFSDHTRDTRATFQGDARFEATFRTGTSFVSATLQSNANPLGVTP
ERDAFESSIFQSDATFVSATFECDAFCFQGVGFERSASGLPLVCAGRVKLSGAVFGCP
VTLSFATRCLECRTWSSTAELRLRYATVDFAHVFVEYFTLTAASADPFVLENGRPV
AEQSLDPIEDIVRIASLRGDAHLVSDVLSGCLFTGTVDLQDRILEGTSFASV
PRETHWRWRPLFTQRRALAEGRFWRASOPTFVPGWNAVAGAHAGAPQALAPVYRA
LRKAFDSNEPGAADFYEGEMEMRHRTGSTRTERGLHGYLLSGYGLRASRALG
WLAAMLVITLLMGFLPKDSKOVATCTVPGDGKVTLEIDKADPQNTGDRFTSE
REFKALTGLNSVVFSSSQDLITAGTYIEMASRVTEPVLGLLAVLAVNRVRK"
  /note="SAV265"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67974.1"
  /db_xref="GI:29603902"
  /translations="MGTRLRLRADFSDCERKLTQKVPVSRPGGVQGLITNTYLSP
TEYDILLASIPAAVLSKTRFSVPLGVDFDGLQGLVLAABEFTTDEQTSFPPSEC
VAEYTDARFTGGRLVETSRQELGLWLAEBYGLRPQ"
  /complement(9955..10614)
  /note="SAV266"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67975.1"
  /db_xref="GI:29603903"
  /translations="MHHVLRRESGCSVLNPLADANRLPVQFRPILLGNHRRRD
DNRFGWTPETLCRLSALGRTAAELSHAI PALAVIGEPARNLQSLSGQAVLRRP
ACRRCMARGFGHGVVTRTPPEHVAQRHQRWLLGDEQVRLHALPELQQANRRHCRIV
RCTRISALAREYAVARDCLRDWFQDGKFSALRQWNRKIDQLGEDPYGDETPVAPPS
G"
  /note="SAV267"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="BAC67976.1"
  /db_xref="GI:29603904"
  /translations="MLEDPAAHGVLDLCTWVHLHETGDEWPATRAHAEFVLPHLREH
RVLVQVARASRLSLEITVIDDSKQQRIVERGPWALWDEYESGGTVPQQGIKCSLH

```

```

CDS
    AKGNWRMPLSPPTTC"
    complement(11530..11979)
    /note="SAV268"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAG67977.1"
    /db_xref="GI:29603905"

Query Match      4.8%; Score 24; DB 1; Length 300425;
Best Local Similarity 100.0%; Pred. NO. 1.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 CTGCCGCGGACGCTGATCTTCGAC 207
Db 203949 CTGCCGCGGACGCTGATCTTCGAC 203926

RESULT 11
AX751873
LOCUS      6378 bp      DNA      linear      PAT 20-JUN-2003
DEFINITION      Sequence 172 from Patent WO03033530.
ACCESSION      AX751873
VERSION      AX751873.1  GI:32134060
KEYWORDS
SOURCE      Mycobacterium tuberculosis
ORGANISM      Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE      1
AUTHORS      James, B., Bacon, J. and March, P.
TITLE      Mycobacterial antigens expressed under high oxygen tension
JOURNAL      Patent: WO 03033530-A 172 24-APR-2003;
Macrobio logical Research Authority (GB)

FEATURES
    source      1..6378
    /organism="Mycobacterium tuberculosis"
    /mol_type="unassigned DNA"
    /db_xref="taxon:1773"

ORIGIN
Query Match      4.6%; Score 23; DB 6; Length 6378;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGCTGATCTTCGACTACCCAC 215
Db 6094 ACGCTGATCTTCGACTACCCAC 6116

RESULT 12
AE007033
LOCUS      18614 bp      DNA      linear      BCT 27-APR-2001
DEFINITION      Mycobacterium tuberculosis CDC1551, section 119 of 280 of the
complete genome.
ACCESSION      AE007033 AE000516
VERSION      AE007033.1  GI:13881331
KEYWORDS
SOURCE      Mycobacterium tuberculosis CDC1551
ORGANISM      Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE      1 (bases 1 to 18614)
AUTHORS      Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE      Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL      Unpublished

```

## REFERENCE

## AUTHORS

## 2 (bases 1 to 18614)

## Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,

## Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,

## Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,

## Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,

## Gill, J., Mikula, A. and Bishai, W.

## Direct Submission

## Submitted (25-APR-2001) The Institute for Genomic Research, 9712

## Medical Center Dr, Rockville, MD 20850, USA

## JOURNAL

## Unpublished

## FEATURES

## Location/Qualifiers

## 1..18614

## /organism="Mycobacterium tuberculosis CDC1551"

## /mol\_type="genomic DNA"

## /strain="CDC1551"

## /db\_xref="taxon:83331"

## /note="clinical strain"

## 71..1483

## /gene="WT1697"

## 71..1483

## /gene="WT1698"

## /note="similar to GP:4378172; identified by sequence

## similarity; putative"

## /codon\_start=1

## /transl\_table=11

## /product="argininosuccinate lyase"

## /protein\_id="AAK45966.1"

## /db\_xref="GI:13881332"

## /translation="MSTNEGLMGGRFAGGSPDAAALSKSTHFDWLAPYDLTASRA

## HTWVLPAGLLTBEQRDLGLDLSLAQVADGSGFLVTDVDSALERGLIDVRGP

## DLGRLRAGSRNDQVAALFRMLRDVRRVATGVLGVGALAEQAQAHPAISMPGKT

## HLQSAQPIILAHHLAHLPLDLDRIVDFDKAAVPSYGSGALASGLGDDPAIA

## ADLFGSAADNSVDAARDFAAEAFVAMIAVDLSRLAEDIIVMSSTGYTHD

## SWSTGSSIMPKQNPDIAGLARGSGRLGNLGLATLKAQPLAYNDLQEDKEPVF

## DSVAQLELLAPRMAGLVASLTENVORMAELAPAGYTLATDLAELVQGVFPRGAHA

## AGAAVRAEORGVLQELTDDELAALISPELTPQVREVLITEGSVSARDCRGTFAPGRV

## AEQLNAIGEAERLRQLVR"

## 1592..2653

## /gene="WT1698"

## 1592..2653

## /gene="WT1698"

## /note="similar to SP:P54157 PID:1256619 GB:AL009126;

## identified by sequence similarity; putative"

## /codon\_start=1

## /transl\_table=11

## /product="chalcone/stilbene synthase family protein"

## /protein\_id="AAK45967.1"

## /db\_xref="GI:13881333"

## /translation="MSVIAVGALPPYRSQRELTDSFVSIPEGVEDIVRQLHAS

## AKVNSRHVLPLEKYPKLTDFGEANKIFIEKAVDLGVQALAGALDESGLRPDLVLI

## TATVTGLAVPSLDARIAGRLRADVRVPLFGLGCVAGAAVARLHDLRGAPGVA

## ALVSVELCSLTYPGYKPTPLGLVGSALFADGAAVAAAGVRAQDAGADPILDSRS

## HLYPDSLRTMGYDVSAGFELVLSRDLAADVVEQLGNDVTTFLASHGLSTTDVGAWVT

## HPGGPKINATETIDLSLSPQALELTWRSGLGNSLSSASVLHLVLEDTIAKPPSPGSG

## LMIAWGPFCFSELVLLRW"

## 2736..9116

## /gene="WT1701"

## 2736..9116

## /gene="WT1701"

## /note="identified by match to PFAM protein family HMM

## PF00550"

## /codon\_start=1

## /transl\_table=11

## /product="polyketide synthase"

## /protein\_id="AAK45968.1"

## /db\_xref="GI:13881334"

## /translation="MNSTPDLVKALRRSLKQNERLKKENRDLIARTTEPVAVVMGC

## RYPGQSPETIWEIWAHRAVDSVEFFADRGDVGALFDDPDPVAGKSYTRCGGLTD

## VAGDAEFFGIAPSEALAMDQORILLVSWEARAGIDITTLRGSTQGVFAGVHG

## SYGQGRVPGDLERYGSLTVASGRVAVVLGQPAVSDVDTACSSLVALLHVAQ

## SLRIGECDDLALVGVTWATPAMFIEFSRQALSADGCKAYAGADCTAFAGAGVL

## VIARLADRUGHVLLVIRGSVNDGASNGLATPNPQAQVITLALASARLGVAD

## VDVVEGHGTGTLDDPBAQILATYQGRPADRLWLGSIKNSITGHTSAAAGVAGVIK

## MVQAMRHGVLPKTLHVDVPTPHVDWSAGAVSLTTEPRFHWVFGFRPRAGVSFGISGT

```

AAATBAASVTEPIAIVGMAACRPPGGVTSADDEWDLISSEQDAIGGFPPTDRGWDLDTL
YDPDPHPGCTYRNGFLYDAGHFAEFGISPREALAMDPOQRLLLETAMETIEHA
GINPHTLHGPTCTGTCNQDVALRVHNAQSTDGFALGTAGSVISGRI5YTFGFE
GPAVSDTACSSSILVHALHQAALRAGECSMALAGVTVVMSPGAFVESRORGLAAD
GHKCA5AADAGTGMGEGVGLIIVLRLSDAHRNGHVLAVRGSANVQDASNGLTAP
NGFSQVRIQOALANAGLSAGDVDAVEAGHTGTLTGDPIEAQALLATYGGDRAGEPL
WLGSVKSNVGTQAAAGVAVIKMVALRHGLLPTLHVDEPSPHVDMSAGAVOLLTE
TPWPGEGEGLRAGTSGVGTNAHVILIEEAPADPGPPAGGAGDAGSDEAAG
SPGVPMVLVSAKSOPALRAQAALHAHLTDHPGLDLADGVYTLAHARAVDRHATLIA
ADRTFLQALQAAGEPHPAVTHSSAPGCTGGEAGKTAFCISGQGTORCPMAWGLI
YHHPVFAALANDICTHLDHLDPLPLLTQNDNDNEDAAALQQTARYAPALPAFQ
VALHRLLDGHITPHYAHENDLSGIEITAAHLGILTLTDATLITQATLMQMPPT
MTLHTPHHTIHTAHENDLAAINTPTSVISGTHPTVQIHTLCOQOGIKTKT
LPTNAPHSPHTNINQLHQTQTLTYHPHTPLITANTPQDQLTTPHYWPTQOART
VDYATTTQTLHOGVTVYIELGPDNTLITTHNLNPNPTTLTLTPHHPHTOHLIT
NLAKTITTHPHYTHDNQPHTHLDTPTYPQHYYWLESQPGAGNVSAAGLDP
TEHPLGATLELATDAGALLAGLSLRSKSHWLDHAGVGVTLISGATFLLELALHAGTY
VGDPRDELITLAVPVPDGGVSQVGAADAGEBRILSVYVARGSSACGGGAGG
VMTCHASGVILVEAAGGVVDGLAGWPPRGAVADVGVDRILAGAGVILVFSGL
RAVNRDGDLLAEVCLPEEAGDAAAGFLHPALIDGVVPLSLPGCTGFBGAGFG
EGVRVPAVGVSVLSRAGVTGVRVRSVAVGGGREGVAVVVGDEAGVPVASVDRLR
RPVDMGOLRAGVSLSRAGVSILYAVOWAEVGPVPCGQAMAHEDVGSGGVPVTV
LRCDAGAGGGGGGGGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
DGPVVVGASVWGLVRSQAQAEHPORFVLLDITDTGTLDTGAGAGWGDGGRVAVV
ACGEPQIAVGERLILAAALXRLSESSGDVPAQRGDTREARRSDVPAQRSGGVPAARSVD
VSGREVLPWLSGGSVLTGTVGLGAARVHLAGVCCVGRDILLVRSRGPDPAGAEGLR
AELALCAEVRIIVACDVGERREVRILLEGVPAGCPLGVVHAAGVTDATIASLTPR
LGVPAKVDAALLDLRLTSGMELSAFVLESSAGIILSGACQGNRYAANAALDALAVR
RRAAGLPGVSLWGLMEASGMTHGLAGTHRRILIRSLGHPMSPTDALALFDAALD
RPVLLPADPAPPLPQLDLPATRRITRTTTTGGANDGAQHLARLAGQTHEQD
LITLALVRSHITAVLGHTEPTDIPDRAFRDGSFDSLTAVELRNRLSRTTGLRPLTTL
AFDHPNPTLTHLITLQLOQPONAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMRWNAPOHTAESADDEKFTSATEAEIKFKIDNDIGLS"
12071. 30790
/ gene="avea2"
12071. 30790
/ gene="avea2"
/ note="contains Ave module 3, module 4, module 5 and
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MOLANEAALKYLEYKRTADLDRTRRLYEVVERSEPIATVGMVA
CRYGGAITSFRLMHLVKSQTDAGEPTDRGMNLEQLYDPRDSRSTSYRSGLYX
DAGDFAFFELSPREALMDPQORLLLETWTFEQGIDPRSMRGSRTGVFGINP
EDYTGTHQPSNAVEGYLLTGSAAISAGRSISYNFLGEGPAITIDTACSSSVLHLL
ACQALRSCECTMALAGASVNAFPVFTEFSRQRLAADGRCKAFSAADATGWSGV
GMLVRLSDARRNGHVLAVRGSANVQDASNGLTAPNGRSOVKVIRQALANHL5
PADVDAVEAGHTGTLTGDPIEAQALVEAGQDRPNRPLMLGLTKSLNIGH5MAAGVG
GVIKMVALRHGLPRTLHVDEPSPHVDSAGAVOLLTETVPWPGGRGLRVSF
GVSGTNAHVILIEEAPNIPSDTPADVPGSAADEADGSGDEAAGSGVWPMVSA
KSQPALRAQAALHAHLTDHPGLDADGVYTLAHARAVDRHATLIAADRTFLQALQ
ALAAEPPHVAITHSSAPGCTGGEAGKTAFCISGQGTQREGMAHGLIHTVPPFAAL
NDICTHLDHLDPLPLLTQDNTQTLTLEAAALQQTARYAPALPAFQVQALHRL
LTDGVHTLTPHYAGHLSGEITAAHLGILTLTDATLITQATLMQMPPTMTLHT
TPPHITLHTAHENDLAAINTPTSVISGTHPTVQIHTLCOQOGIKTKTPTNHA
FHSHPFLVNLQHLQTLTYHPHTPLITANTPQDQLTTPHYWPTQOARTVYAT
TQTLHOGVTVYIELGPDNTLITTHNLNPNPTTLTLTPHHPHTOHLITNAKT
TTPHPHYTHDNQPHTHLDTPTYPQHYYWLESQPGAGNVSAAGLTPST
PLRVLVKSPHQRDEELLRLVTHAAVGHVATPEVVPNKAFGLGFDLSAIAQLR
NRLADVDPATLIPDPTPMALCOFLRAIVGADTGTTLPLTAPVADPEPIAV
MACRYPGDVRTVDDLQWVSGGDAIGGFTNRGWDLDTLXNPDHHRGTSYRSFG
FLYDAGNPDPDFGSPREALAMDPOQRLLLETAWESLEHACIPDSLRGTSYRSFG
LTYHDYARPTAPAGEGYLGHSAGSIASGRVAYALGLEGPALTVDTACSSSVLAL
HLAQALRSFGCSMALAGVTVMSPTAGVFSRQRLAADGRCKAFSAADATGWE
GVGMLVRLSDARRNGHVLAVRGSANVQDASNGLTAPNGRSOVKVIRQALANAD
LTPADVDAVEAGHTGTLTGDPIEAQALVATYQDRPGNEPLWLSGM5NIC5HAAG
VGVKIKMVALRHGLPRTLHVDEPSPHVDSAGAVOLLTETVPWPGGRGLRAGVS
SFGVSGTNAHVILIEEAPNIPSDTPADDAFGEAADDPVGEAAGDAGTGEATGPA
AGSGFVWPMVLVSAKSQPALRAQAALHAHLTDHPGLDADGVYTLAHARAVDRHATL

```

gene

CDS

Query Match 4.6%; Score 23; DB 1; Length 64957;

Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 ACCTGATCTTCGACTACCCAC 215

Db 60213 ACCTGATCTTCGACTACCCAC 60191

RESULT 16

AF521085/c

LOCUS AF521085 132544 bp DNA linear BCT 23-MAY-2003  
DEFINITION Streptomyces nanchangensis NS3226 polyether ionophore nanchangmycin  
biosynthetic gene cluster, complete sequence.

ACCESSION AF521085

VERSION AF521085.1 GI:31044135

KEYWORDS Streptomyces nanchangensis

SOURCE Streptomyces nanchangensis

ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE 1 (bases 1 to 132544)  
Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.

A Complete Gene Cluster from Streptomyces nanchangensis NS3226

Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin

JOURNAL Chem. Biol. 10, 431-441 (2003)  
REFERENCE 2 (bases 1 to 132544)  
AUTHORS Sun, Y., Zhou, X., Dong, H., Tu, G., Wang, M., Wang, B. and Deng, Z.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-2002) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

FEATURES  
source  
1. .132544  
/organism="Streptomyces nanchangensis"  
/mol\_type="genomic DNA"  
/strain="NS3226"  
/db\_xref="taxon:204925"  
complement (682. .2052)  
/gene="nanT1"  
complement (682. .2052)  
/gene="nanT1"  
/notes="putative integral membrane transport protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanT1"  
/protein\_id="AAP42848.1"  
/db\_xref="GI:31044136"  
/translation="MHDITKDDPRASRRSGVRFAVASLTGTALDWDYDPAVTSAAAA  
LVFGDLFFPSEDPLTGLTFLAFSTVAVGYSRPLGGIVFGRIGDVI GRKKVLIATLVLI  
GAATFLIGLPTTSTIGVAAPIALVLRFAQGVGGEWGGVALLSEFDSRRRGFY  
ASAAQVGPAPNGLANGVLAALGALLTEAQFEAWGWRVAFLLSGALVAFGLIRAKLE  
ETVFKAMEAQSRPRAPIREVFTPTOPRALLAILCRVGPDLVYAMFTVPVLTYATGE  
LGMRSALANVLGSLQVFLMPLAGVSDRINRVLXGCAVAGWPEFLPEPMIG  
GGSWIPLAGVVLVSHLSTIGPQAFIAEQFSRPLRYTGSSLATYLAGIIGATAP  
LFTTLLSAYGSWLPIALYIAVAAVSLAGVLIGRDEPAEAVSDAQLATPAQAQTADA  
ARPTAV"  
2064. .2879  
/gene="nanT2"  
2064. .2879  
/gene="nanT2"  
/notes="putative ABC transporter"  
/codon\_start=1  
/transl\_table=11  
/product="NanT2"  
/protein\_id="AAP42849.1"  
/db\_xref="GI:31044137"  
/translation="MLGHRPSPGILLVYRSYAQAKRKLPAEKSTTADRHNPSSAPM  
ARLTFELPGDSTREVDIVQINAGYAGRSQDDVAATHAEALGELGVPPTPALYPVA  
PYLAQOQDRVAVOHRRTSGAEAWLVAGDGLLITAACTDTRDLGVHGVAMSKNAG  
PDVLARAWLADVEPLDILTLFAWVTRDGTETEOHGTLEALLTPAYVVDVLSRD  
ALTPGVLIISGTPMTFQVQFADTWVVELGDPATGTIRLAYDVHWPPEIG"  
3117. .3818  
/gene="nanT3"  
3117. .3818  
/gene="nanT3"  
/notes="putative two-component response regulator"  
/codon\_start=1  
/transl\_table=11  
/product="NanT3"  
/protein\_id="AAP42850.1"  
/db\_xref="GI:31044138"  
/translation="MSTRCTGSLNMGWVSVELGPLSVLVDHPVVRAGISAWCAAA  
DAPISVVAEGANVSVALHGPGRGADVVMILLQNGRPAYDELQELVAQERKVVVTM  
RDSQDAALTCDMGSAITITKAGQRHLVKAIRAAABDIPTTPPSLAGAFGSDTRQSR  
FVLSREVVLVEWFQESKAVVAQSLGISERTVTYLDRIKRYANAGRPAATTKAKL  
VARAVQDLIALDEL"  
3174. .4923  
/gene="nanT4"  
complement (3724. .4923)  
/gene="nanT4"  
/notes="hypothetical chemoreceptor"  
/codon\_start=1  
/transl\_table=11  
/product="NanT4"  
/protein\_id="AAP42851.1"  
/db\_xref="GI:31044139"  
/translation="WGEDFRQLITRHQIRGIRAGGLVILVTIHTIGSALPGLLGLGEY

gene  
complement (6768. .7493)  
/gene="nanR1"  
complement (6768. .7493)  
/gene="nanR1"  
/notes="putative regulatory protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanR1"  
/protein\_id="AAP42853.1"  
/db\_xref="GI:31044141"  
/translation="MSPLRVVDEKGGSPLEAQLAAVLAVLLIRSDOVVMAEQLTREI  
WGEPLPRDTSIYVICSLRKLRLSPGQONPILITRPGYMLRRRSECEDFVQQLV  
ADGKCVRYGEPAQALCFEQALALWRGPAGNVQTPGPIVAGFLKRUDEMRIBCAEWS  
AEALHGRHRELVLGRLSRSLVAEFLFRETTFQRLMALYRSQKSEALGIYESARRVL  
DRELGLPGKRLRELHRTVLAST"  
complement (7573. .8334)  
/gene="nanR2"  
complement (7573. .8334)  
/gene="nanR2"  
/notes="putative regulatory protein"  
/codon\_start=1  
/transl\_table=11  
/product="NanR2"  
/protein\_id="AAP42854.1"  
/db\_xref="GI:31044142"  
/translation="MRYEILGPLRVVDRSSVTSISKRIEVLAAILVRADEVVSSD  
LVKIMGTRAPROAAALHVVSQRLKFLRRPNRTESPIVTRSGYVLRGLDDELDFH  
VFQGLVROGAAQAGRTHEACAYBEGLEFWRGVLELDLGDALVNGFVCWAQOARL  
ECVSEFVAGLELGRHREFVSYLYARLEELPLNEAFYQLMALYRSGCRADALHVVQ  
SARAVLSEELGLEFSAALKRHHAVLVADELEAA"  
8919. .17627  
/gene="nanA1"  
8919. .17627  
/gene="nanA1"  
/notes="PKS"  
/codon\_start=1  
/transl\_table=11  
/product="NanA1"  
/protein\_id="AAP42855.1"  
/db\_xref="GI:31044143"  
/translation="MRFYGVWAGSSPSHAQQTSPVAIVGLACRLPGAPDPFAFWR  
LLRAGENAVVPPDSRLTPEFSPPIFAGLEHVDITFDAGFFGLSPREATMPPQRL  
MLELAMEDAGLGPKNLAERTAVFTGAIWDDYATLLHRRRNDIAITHTWAGLH  
RGLIANRVSHLLGLRSLTVDAQSSGLVAVHLACSLRRGEADLALAGGVNII LAE  
ESMRMAEQFQGLSPDCRCVTFDARANGFVRGEGGMVLLKPLAAAVADGDPVVCVE  
GSVNNNGATDGLTRPSADAQTDVVRQMQRAVSPAELVOYVLEHGTGTGTPVGDPIEA  
ALGALGDADGCHRDRAVDAPLRVGSVKINVGHELEAAAGIAGLKTALSIIHRLPFP  
SLNPAITNPGLPLAELGLRVQTAFGPWPDPERRRLTAGVSSFGMGTCNCHVVVIAEFPAP

AVLPDRPTAPGDSARTAPVMPVWSAASFKAALTAQAAALYEHRAHPLGLHFVDVIGH  
ALATRTAPPHRAVLRSDDELVDLDALESTQTSVIRGRAGGRVAFPLNRTGAY  
ORPMGRELPAAYVPFADALREYCAHLDDMLDDTDLDDIMFAEAPDGPPLNRTGAY  
TOPALFAIEVALYELTSMGVTPOHLMGHVSGBITAAHVAVLSLDPDCTVLAARGEL  
MQSITAPGAWAQTAABEQALEYGGVRGLAAVNPASVVISGDRNAVAAATRAAM  
RARGKATVLKVSASHPHDGLDGLTRVAGLTAAPAIFVNSNLTGGAATEAQL  
RSPDYWADHARQAVFEDAGVHLCDAAGVDFLELDPDASLTGMARSAAAWADAPRP  
VAVAVORGRPEAQSFVVSMAQAARHGVGDVMAAFAGHETTRAPLPTAFQDRHWP  
DGLDERGARPSFVPSVTRPVPVADPADRAASPELLALVRTHAALVLGHNSP  
DGDIPALTFQGLDFSLAATLSERLSAATTELPATLTEDHPTNPAVLAARAHG  
OPTAPATATQMAEDPVAIVAVSCRYPGVSGESGALMRLLDEGVDAVCEFPQDRGW  
DLAEFORAPGSGSATGGGLYAGDFDEAFEGISPREALAMDPOQRILLSELWE  
LLERAGTPPASAGSATGVVGTAVDYGPRHLEAEHLLTGSTPVSAGVAV  
ALGLEPAPLADTACSSSLVAMHLAQLRQECDCALAGGVMTMATPGTFEFSSQR  
GLAPDGRCKPFAAADCTGMSAGLVLRLERLSDARRNGHQLVIRGSVAVNDGASN  
GLSAPNGSPQSVTRQALANARLEPADVDAVEAGTCTTLGDPLEAALLATYGGOTT  
DLRPLWLSGSKNIGTHQAAGVAVGLKYMALRHGLPLASLHIDAPSPHLDMSDCTV  
RLISEPVDWPGTDPGSDRRRAVSVFSGISGTHNAHLLEQADPHPEPTEPSGGVVP  
WLSARTADALRAQAAGLAEWVTAGAPRSPASPSPSPADVQSGWGMVGLLDSPPVF  
VSGTDRDELLSGLRAVDAGLAPAAVSAAGAPGPMVPPGQSGWGMVGLLDSPPVF  
AARMAACEAALGFEDVMSLTVLRGAPGAPGPRVDVLQCLWAVMVSLAAVWESYGV  
TPTAVVCHSGEIAAACVAGLSLRDGCARVVALRSOALRALAGHTWASIALSGABAE  
REFLADLGAARVTVAVFNGPYSVSGPTDQVAVVAAACEAGHRARTIDVYASHG  
PQVRLADTIRTDLADUSPGASDAVFSVGTARQPTDELDADYFTNLKQPVRFASA  
IDALLAAGYRVFIEVSPHVLIPALRECFEEAEVAAATVTLRRDQGGPDQVAFALGD  
GVFAGLAVDSRFVGVGREGAGDGBHRTVELTPTFPQRRYWLAPDHGREGRTAG  
VGTREPAGHLLSVELADGGLVSGRLPGDAWVGHTVAGVQLVPGAVLWDALLIA  
ADPAGASLELLRLAPLESGLSPGSLSEPSAGVLAOVAVGAPDSEGRERLISRPAD  
AGAGEWHTCHAVGSLAPGPPAPADTGTATVPFPAGAEALDPAGLYERARRGYGV  
PALRGVVALWGDADLVALPEEAGGEGGADGGTAGFGLHPHVLDDAALQPALU  
AEPDGTGEGAGPEARLWLPFMSGVRLMATGAPARVRLSPLDGGGVDADERELLI  
EVSDDTCAVLSVAVLRVTRTVQVREASGAAGGLFALDWTVPAPQEPSGADDDAG  
CVALGEPAPTEPGVDCRDTVDLPALLAALDAGAPLPVVMWRPPAADPGAPEDDQ  
LSAVRGVAALRAWVAEPRLTVSLAVTVTRGAVAGGEGEPVDLAAAGGARGVQ  
AEPHDRVILVDDVDVMDGADTDDICAAAGLAAALGEPQVALRGDILLAPLARSAA  
TPGVAVPNGTVLDSVGGSLAGSVAEHLVRAGVRLHLLVREFGADGAYDITDROQ  
AQHVMTVPRDRTAALERVVAQVDPHAPLTVGVHVSADSIETSGAARGWAVAGV  
RALHQAADALPSVRFVLTDAATAGDGAAPERAAGACFAAVTVDRVRAGLGLDVA  
FGWAAADDDGAGDGGRTGVLGADGLALRAACRADRPRLVADIRTRALTAHPA  
HELPAALRTIGASASAGGAPFVRVVAAPGRTTDSWSELVGLGPAERRAVIELV

Query Match 4.6%; Score 23; DB 1; Length 132544;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 ACGCTGATCTTCGACTACCCAC 215  
Db 95473 ACGCTGATCTTCGACTACCCAC 95451

RESULT 17  
BX248339  
LOCUS  
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 6/14.  
ACCESSION BX248339  
VERSION BX248339.1 GI:31618223  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis subsp. bovis AF2122/97  
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1  
AUTHORS Garnier, T., Biglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Iacoi, C., Monsemp, C., Simon, S., Harris, B., Akkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. and Hewinson, G.  
The complete genome sequence of Mycobacterium bovis  
Online Publication  
PNAS 10.1073/pnas.1130426100 ( Microbiology )  
REFERENCE 2 ( bases 1 to 300050 )

AUTHORS  
TITLE  
JOURNAL

Garnier, T.  
Direct Submission  
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28.rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium Bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NP, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
source

1. 300050  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
165. 398  
/genes="secG"  
165. 398  
/locus\_tag="Mb1475"  
/gene="secG"  
/locus\_tag="Mb1475"  
/notes="Mb1475", secG, len: 77 aa. Equivalent to Rv1440,  
len: 77 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100% identity in 77 aa overlap). Probable protein-export  
membrane protein secG, similar to many e.g.  
P93888[SECY MYCLE PROBABLE PROTEIN-EXPORT MEMBRANE (77  
aa), FASTA scores: opt: 450, E(): 6.7e-24, (96.1% identity  
in 77 aa overlap). Start changed since original submission  
(-40 aa). PART OF THE PROKARYOTIC PROTEIN TRANSLATION  
APPARATUS WHICH COMPRISE SECA[Rv3240C, SECY[Rv2587C,  
SECZ[Rv0638, SECF[Rv2586C, SECG AND SECY[Rv0732."]  
/codon\_start=1  
/transl\_table=11  
/product="Probable protein-export membrane protein secG"  
/protein\_id="CAD96142.1"  
/db\_xref="GI:31618224"

gene

STVVERNLDRITLFTVGTGILVSLVIGVALLKYR  
complement(537..2012)  
/gene="PE\_PGRS26"  
/locus\_tag="Mb1476c"  
complement(537..2012)  
/gene="PE\_PGRS26"  
/locus\_tag="Mb1476c"  
/notes="Mb1476c, PE\_PGRS26, len: 491 aa. Equivalent to  
Rv1441C, len: 491 aa, from Mycobacterium tuberculosis  
strain H37Rv, (99.6% identity in 491 aa overlap). Member  
of the Mycobacterium tuberculosis PE family, PGRS  
subfamily of gly-rich proteins, similar to  
YODP MYCTU[O50615 hypothetical glycine-rich 40.8 kd  
protein (498 aa), fasta scores: opt: 1625, E(): 0, (55.2%  
identity in 518 aa overlap)."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="PE-PGRS FAMILY PROTEIN"  
/protein\_id="CAD96143.1"  
/db\_xref="GI:31618225"

CDS

2119..4419  
/gene="bisc"  
/locus\_tag="Mb1477"

gene

2119..4419  
/gene="bisc"  
/locus\_tag="Mb1477"

CDS

2119..4419  
/gene="bisc"  
/locus\_tag="Mb1477"

gene

2119..4419  
/gene="bisc"  
/locus\_tag="Mb1477"

CDS

```
2119..4419
/gene="bisc"
/locus tag="Mb1477"
/EC number="1.1.1.1"
/notes="Mb1477, bisc, len: 766 aa. Equivalent to Rv1442,
len: 766 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 766 aa overlap). Probable bisc, Biotin
sulfoxide reductase (EC 1.1.1.1), similar to
BISC EC01|p20099 biotin sulfoxide reductase from
Escherichia coli (739 aa), FASTA scores: opt: 1271, E(0.0,
40.2% identity in 744 aa overlap)."
/codon_start=1
/transl_table=11
/product="PROBABLE BIOTIN SULFOXIDE REDUCTASE BISC (BDS
reductase) (BISO reductase)"
/protein_id="CAD96144.1"
/db_xref="GI:31618226"
/translation="MQVYTSATHVGTARVHGSDIAVAALASDTPNPAPQLNPLGA
VHRGRIRANPVRRLQHPGSSARGAEFEVSVDELIELASELRVTDYRNE
AIYGSYGNASAGRHFAQVHRLNMLGGVYTSRHSYSAGASEVIFPHIVGAALFE
ALAEITTVDIIVDHTALLVAFGLPVKNTAVMGSTTAHPDRDYVGRYARGRLVSV
SPRLDDIAALGDLDDCHWLAAPVCTDVAIMGLAVLATSLADRAFLGRCYCTGYE
RFRVLLGLDDGIPKTPPEWAALSGLAAGDLRLARMAEHRITLITISLSQRIERGE
QTVWAAIAGMQIGLGGGFGHGYSSNGVGNPPCLGCPALPGQNNPVSFTIPVA
AISELLQPGQRILVYNGRLLELPDIKCVYAGNPNPFRHQNLPRLRRALSVDITVYH
EYWTAMAKQADIVVTTTSFERDDPAASKNTPLIAMPVPPYANARDVHTFSAL
AHLRIGKQTEGERSAREHEMDYKWSAELDPVPSFAEFTWRTGLELPTRTGLTWL
ADPRADPAHLPTGTPSGRTEIFSDTVDAFPALDPCAGHTWYBSEWLGPPRAARYPLH
LIANQPTRLHSLDGHGASMASKIRGSRPIRHDPDAAARELTDSGDIVRVENDRGAC
LAGVVIDGRLFRKVLQSLGADFPADPRDPSMCMVGNPNALNSDGSSTSLAHGSTG
QHVVLIDGRTGLPVRHAHEPRLA"
```

gene

CDS

```
/locus tag="Mb1478c"
/notes="Mb1478c, -, len: 161 aa. Equivalent to Rv1443c,
len: 161 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 161 aa overlap). Hypothetical unknown
protein."
/codon_start=1
/transl_table=11
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAD96145.1"
/db_xref="GI:31618227"
/translation="WGVYAEPLVIRQSVVAAPAOVWQVVTPEGINDELPRMTWS
VPRGAGMTVDVTPICAPILGRMLRGLVLPDYDLSIAELPGRFEDSTLSNR
G"QHQHETVPEGTIKIVDRITPQTRAGLRFAAPLIAAGLRALFGHRRRLQRHFAQ
G"
```

gene

CDS

```
complement(5634..6044)
/locus tag="Mb1479c"
/locus tag="Mb1479c"
/locus tag="Mb1479c"
/notes="Mb1479c, -, len: 136 aa. Equivalent to Rv1444c,
len: 136 aa, from Mycobacterium tuberculosis strain H37Rv,
(100% identity in 136 aa overlap). Hypothetical unknown
protein."
/codon_start=1
/transl_table=11
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAD96146.1"
/db_xref="GI:31618228"
/translations="MTVMADRSRGPAPVRMRMTKTLQAAALNADKTVQVBDVLDGLGK
TMAELNSSLQSLNSTRVERDGLHLEGLTHSLDGLAKRLILVPEVPAIVDRIDIVY
SLGETVMSPLSVTEHVAVRGVLDRLNRVTVEPTN"
```

gene

CDS

```
/gene="devb"
/locus tag="Mb1480c"
complement(6061..6804)
/notes="devb"
/locus tag="Mb1480c"
/EC number="3.1.1.31"
/notes="Mb1480c, devb, len: 247 aa. Equivalent to Rv1445c,
len: 247 aa, from Mycobacterium tuberculosis strain H37Rv,
```

```
(100% identity in 247 aa overlap). Possible devb (PGL),
6-phosphogluconolactonase (EC 3.1.1.31), belongs to a
different family to the upstream gene zwf2. Similar to
e.g. DEVB_ANASP|p46016 putative glucose-6-phosphate
1-dehydrogenase (239 aa), FASTA scores: opt: 439, E(0.0,
2.6e-20, (34.0% identity in 247 aa overlap). BELONGS TO
THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE ISOMERASE
FAMILY. 6-PHOSPHOGLUCONOLACTONASE SUBFAMILY."
```

```
/codon_start=1
/transl_table=11
/product="PROBABLE 6-PHOSPHOGLUCONOLACTONASE DEVB (6PGL)"
/protein_id="CAD96147.1"
/db_xref="GI:31618229"
/translation="MSSSIEIFPDSLDILVAAGKRLVGAIGAAVARGOALIVLTGGG
NGIALRLYLSAQAOQIEWSKVLFWGDERVVPEDDNLKQARRALLNHVDTPSNQV
HPMAASDGGDGLDAAALAYQVLAASAPGDPAPNEDVHLGMCPEGHINSLFPHS
PAVLSTRVAVVADSPKPPPRITITLPAIQSRREVLLVSGFGKADAVAAALGAD
PVSVPAAAGVGRQNTLWLLDRDAAKLPS"
```

gene

CDS

```
/gene="opca"
/locus tag="Mb1481c"
complement(6801..7712)
/genes="opca"
/locus tag="Mb1481c"
/notes="Mb1481c, opca, len: 303 aa. Equivalent to Rv1446c,
len: 303 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.7% identity in 303 aa overlap). Putative opca, OxpP
cycle protein. Highly similar to S72774 B1496.F1.30
protein from Mycobacterium leprae (265 aa), FASTA scores:
opt: 1056, E(0.0, (70.3% identity in 239 aa overlap).
Also similar to OPCA_NOSS2|P48971 putative oxppcycle
protein opca from Nostoc punctiforme (465 aa), fasta
scores: opt: 177, E(0.0, 7.3e-05, (23.4% identity in 321 aa
```

```
Query Match 4.6%; Score 23; DB 1; Length 300050;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 193 ACCTGATCTTCGACTACCCAC 215

Db 252783 ACCTGATCTTCGACTACCCAC 252805

RESULT 18

AP005024/c 302675 bp DNA linear BCT 10-MAY-2003

LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 4/30.

DEFINITION AP005024 BA000030

VERSION AP005024.1 GI:29604389

KEYWORDS Streptomyces avermitilis MA-4680

SOURCE Streptomyces avermitilis MA-4680

ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE 1

AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,

Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,

Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.

Genome sequence of an industrial microorganism Streptomyces

avermitilis: deducing the ability of producing secondary

metabolites

Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)

MEDLINE 21477403

PUBMED 11572948

REFERENCE 2

AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,

Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.

Complete genome sequence and comparative analysis of the industrial

microorganism Streptomyces avermitilis

Nat. Biotechnol. 21 (5), 526-531 (2003)

MEDLINE 22608306

PUBMED 12692562



```

3 (bases 1 to 302675)
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinoe,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
Kushida,N., Director-General of Biotechnology Center, Shiba,T.,
Sasaki,Y. and Hattori,M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:bioente.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun
Ishikawa(*2), Akihito Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
Shinoe(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.ls.kitasato-u.ac.jp.
Location/Qualifiers
1. 302675
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/notes="This strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermitilinus"
12. 425
/notes="SAV750"
/codon_start=1
/transl_table=1
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68460.1"
/db_xref="GI:29604390"
/translation="MARIICLRPRTAKAQLSLAGQGVARNLDAVALDFDSGG
PAMGLANTRPAALRIREALVDGNPGFTAOEALTGYLKAERGLGRVAETVDRPTVA
LAIVGTAHLHMTSMFGADPRPHMARLUVALVDS"
1504..2592
/notes="SAV751"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC68461.1"
/db_xref="GI:29604391"
/translation="MRTVVEEELLVDPETGPKALSTAVLARAEQVDPDQVPEKE
LHYOMAEQYGIAMOEQLTCGCHVAVVESDEEGVAVVDRIPLPLVALSANSFPFW
RQYOMAEQYGIAMOEQLTCGCHVAVVESDEEGVAVVDRIPLPLVALSANSFPFW
QCRDSSYSRSRWGRNPSAGPTFLFSGPERYHROQVADWATGVLLDDGMVYEDARL
SHRYTVIRVADVCLHEDTQALITATLARGLVESAARDWRAGREPEDSVLLRLAAW
QARAGLSGELLHPVTMRLEASVVRALLHVGDALADAGDLDAHEAVAEVLEERGN
GARVORELLERTGSLRDVTVECVRTQG"
complement(3011..3400)
/notes="SAV752"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC68462.1"
/db_xref="GI:29604392"
/translation="MQITASTVSLTVADVAASQAFLTALHGYTVQAAADGFASLIRAD
AVDIVLARGTVLPEQQRDRHAGLLIATFLADGLHQEKRLREAGVEITWPLREEP
WGERLFIQITDPNGVIIQFVEMAYSDDT"
complement(3457..3648)
/notes="SAV753"

```

---

```

/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC68463.1"
/db_xref="GI:29604393"
/translation="MFELDTTTSALTIGLGEVPAGDVDPAPRGYDLRLAADRA
VGRGRFRVVRNIIPRRRGY"
3755..4408
/notes="SAV754"
/codon_start=1
/transl_table=1
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68464.1"
/db_xref="GI:29604394"
/translation="MGTEITSPQPVGLRSEKHEHETQLISDHATRLFLAQGFQTTI
AETAAARMKATVTVYFARKDLALDHQEEFTASLANAVADRRPGSALAALIRAPH
DSLATQNPVAGFSGPAFARMVADSPILTCLRGLDHQREHLAQALADATDTTDDTIT
VHTAAGLLGTVRHVLFORIQDLTLAGHPNDRIHTITREATQAFDLLEPALAHYATA"
4587..5627
/notes="SAV755"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="BAC68465.1"
/db_xref="GI:29604395"
/translation="MTSPKTRTTTSPVACPATGSPSSQATSPAPSVPAVTRACSTWT
CRGRDIRPAAPVPRSQADPEADFPALVSRTLSDLTGCRHVLRSVSLDSFDVP
LVTAAGTAHDAPALRLTERPFVRENGTCRGAARPAPPGAHPRDWRVSGRPASDGP
APTSRRPAASPAQPNQHRGSANVTMEATPKWLAVALVSLDVAHHTGPAALRHLS
AAGSNVSGAGGIDGDIYAMAYNAAGAVNLTAMALDHAPEFVRVNHAGHFGFTLHSG
ITAGATPDTVRSFGEIERIPLGRVAEPEIASVVAFLAGDPALDYRNSNPGRRPHRC
YRAAPHRDVNAS"
complement(6101..8071)
/notes="SAV756"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC68466.1"
/db_xref="GI:29604396"
/translation="MERSKDRPVSLQDRARYFDRTLARSTGLMGVLVILCLAVV
PVSLIIVTDPGSPRSLSGRLAAVWRSATLRLGTGTGTLRLMLLSALLGLVALLCV
STLVGTVITGLAERLBSRGSTVLEQGHAVLGSDQVSTVVGELVVAOASSYRPA
VVLAEADKTEMERALAAHVGPAQRTLCVRSQSPADPGLVALVSPQSTVILVPSG
EPTADRVLLALRAVLGEGTGPVLAALDDRYAPARLARGRTVLETDTVT
ARLIAQCVRGELSLVRLDLDFAGDFHIAEATAFHGPGFCAALLGHATSCVVGILT
AEGSTLNPAAATIVAPGSLVLTDRDSDARPEDCRHLVPEPSALMAQPPEDAAHL
LLGNWRRAPLVYNQLRTARPAGSLVDVTDVAVGPTREPESGAARLAVRFQADLS
RPETLLGLDLPYDGLVIGPEPGDGRDRLTLLALLLEERTGRMRVITEL
TDDRNRPLAPVNGSLVIVSGELTGLLLAQIAQRHRLAPVDFELFSADGSTICLRQAG
HYVRTGCTTPTATVVAARDRGECAIGYRHRDRAIPPDHGIRLNPDKSERRWSADD
QVWVATEPTPAPAAGQDTBEHQGPADTRA"
9086..10072
/notes="SAV757"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC68467.1"
/db_xref="GI:29604397"
/translation="WGRKKGIGIALNTVMLVGCASCGGSGDKANGDDAKASQKPG
ASAPSPKPPAGPMLLETITPQTGTTVGVMPIVVTNPAKARAAVEKHMVSA
SQPVSGAWHVGDKRADWRPKTYWPSGFKVIDAMKGVNNGNGRIVGHSYHTFKIG
DVRADVSVGTGTMKVTNGKVRVRLTSINAGSAQYPTWGTMAVIDKQKRVHTSCSV
GISCKGSPNYDLTLPMDWHLTQSTGYVHYSTGDPTEPGSARGSGHCVHLSMDAK
WFGYQVKGQDPVTITGSPRAKADNGYADPNLGDWDLQWLAGSASGEKTTATL"
complement(10889..11494)
/notes="SAV758"
/codon_start=1
/transl_table=1
/product="putative acetyltransferase"
/protein_id="BAC68468.1"
/db_xref="GI:29604398"
/translation="MQRAVPRRAIMAVMARPQIIECGDFVLRSGRQDDFAPAFKLE
ESLDHLREWEPPWAFHSEESTDFLAKSESKWANGDVVNYALAKDGTILGMCQTYRGA

```



EPOGRRIGWHLHPAATGOGIATRAATAAMVTEMFTLPDVEYLEVAHDLANTSSAALPRR  
 LGFTVREQVTEPPAAGSGGIDVVWRLNRPTPPVDGSRRP"  
 11580..12161  
 /note="SAV759"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC68469.1"  
 /db\_xref="GI:29604399"  
 /translation="MRSRAPTRCWPGSTRPSDATALCWRNWRWTCGTGTGPRPRDRI  
 SAACFFRAVPPALLQAPGGKILLTGLWLYGARVLLTVADDDTAEGQLRGDTVSF  
 MSARHAPTYGNPAHATVPEKATRTTRHSPEIRITATDEALHLRFLAQSAVSDAQM  
 TTVGDVHLVDVVTGSRDPHPWRGQLGCTGRG"  
 complement(12097..12768)  
 /note="SAV760"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative multicopper oxidase"  
 /protein\_id="BAC68470.1"  
 /db\_xref="GI:29604400"  
 /translation="MQAFSPHGFHAEDYATLPSRAKNEAIYGVNDRSCMLAYHD  
 HMGSMALNVAGLGLIDVDPDERLGLPGAEFVLLIRDRITFNQDGLAYNTA  
 REGDTPVNGAYFLAVPEPRYLRLINASNEPFWRLRFDVPRDVLQPPFLWLI  
 GTDGGFRAPLKLMDFLISSAERYDLIVDFSGMPRTSRCHSFTGLAYSPTVLARGE  
 GRGSG"  
 complement(12621..12968)  
 CDS  
 Query Match 4.6%; Score 23; DB 1; Length 302675;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 193 ACGCTGATCTCGACTACCCAC 215  
 |||||  
 DB 297929 ACGCTGATCTCGACTACCCAC 297907  
 |||||  
 RESULT 19  
 BX842577  
 LOCUS BX842577 347496 bp DNA circular BCT 21-NOV-2003  
 Mycobacterium tuberculosis H37Rv complete genome; segment 6/13.  
 ACCESSION BX842577 AL022000 AL022001 AL022002 AL022003 AL022021 AL123456  
 274020 278020 281360 285982 295554 295586 295617 295890 296073  
 297345 298268 298943  
 VERSION BX842577.1 GI:38684030  
 complete genome.  
 SOURCE Mycobacterium tuberculosis H37Rv  
 ORGANISM Mycobacterium tuberculosis H37Rv  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 REFERENCE 1  
 Cole, S.T., Broesch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
 Barrell, B.G.  
 Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence  
 NATURE 393 (6695), 537-544 (1998)  
 JOURNAL MEDLINE 98295987  
 PUBMED 9634230  
 REFERENCE 2  
 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.  
 Re-annotation of the genome sequence of Mycobacterium tuberculosis  
 H37Rv  
 MICROBIOLOGY (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)  
 JOURNAL MEDLINE 2225591  
 PUBMED 12368430  
 REFERENCE 3  
 (bases 1 to 347496)

Parkhill, J.  
 Direct Submission  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium  
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome  
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique  
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 On or before Dec 4, 2003 this sequence version replaced gi:3261541,  
 gi:3261543, gi:3261544, gi:3261547, gi:3250699, gi:3261584,  
 gi:3261625, gi:3261654, gi:3261718, gi:3261771, gi:3261785,  
 gi:3242249, gi:3242245, gi:3242257, gi:3261824, gi:3261839,  
 gi:3261842.  
 Notes:  
 Details of M. tuberculosis sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/).  
 FEATURES  
 source  
 1..347496  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 116..1219  
 /gene="adh"  
 /locus\_tag="RV1530"  
 116..1219  
 /gene="adh"  
 /locus\_tag="RV1530"  
 /EC\_number="1.1.1.1"  
 /function="Catalyzes the reversible oxidation of ethanol  
 to acetaldehyde with the concomitant reduction of NAD"  
 /note="RV1530, (MTV045.04), len: 367 aa. Probable adh,  
 alcohol dehydrogenase (EC 1.1.1.1), zinc-dependent,  
 similar to many e.g. AE0009|AE000958|23 Archaeoglobus  
 fulgidus section 1 (402 aa), FASTA scores: Opt: 423, E():  
 1.8e-19, (31.7% identity in 341 aa overlap). Contains  
 PS00059 Zinc-containing alcohol dehydrogenases signature.  
 TBparse score is 0.919."  
 /codon\_start=1  
 /transl\_table=11  
 /product="Probable alcohol dehydrogenase adh"  
 /protein\_id="CAAL7590.1"  
 /db\_xref="GI:2916953"  
 /translation="MSDGAIVRALVLEAPRLVRQYRLPRIGDDALVRVACGLCG  
 TDHGYTGELAGFAFVPGHETVGTAAIGPRAEORWGVSGADRVAVEVFOSCROCAN  
 CRGSEYRCVRHGLADMVGFIVDREPGLWGYAEVQYLAIPDSMLRVAGDLSPEVAT  
 LFNPIGAGIRWGVITPETKPGDVVAIVIGPGICLCAAAKAGAGFVMTGLGPRDA  
 DLALAAQFADLAVDAIDDDVAALBTQGLADVVDVDTAKAPAFQAALAKRPA  
 GTVVAGTRGVGSGAPGFSFEDVYVFKELRVGLGADVATAYRAALDLLVSGRYPFASL  
 PRRCVRLGAEDLLATMAGERDGVPPIHGVLTP"  
 302..346  
 /gene="adh"  
 /locus\_tag="RV1530"  
 /note="PS00059 Zinc-containing alcohol dehydrogenases  
 signature"  
 1216..1782  
 /locus\_tag="RV1531"  
 1216..1782  
 /locus\_tag="RV1531"  
 /function="UNKNOWN"  
 /note="RV1531, (MTV045.05), len: 188 aa. Conserved  
 hypothetical protein, similar to RV0464c|MTV038.08c (190  
 aa), FASTA scores: E(): 4.8e-10, (30.9% identity in 175 aa  
 overlap)."  
 /codon\_start=1  
 /transl\_table=11  
 /product="CONSERVED HYPOTHETICAL PROTEIN"  
 /protein\_id="CAAL7591.1"  
 /db\_xref="GI:2916954"  
 /translation="MTTSRVPLLPVDEAKAAADACGVPDYMALSTIFQVLLNHPRLAR  
 TFNDLLATMLWHTLDSRLRLVIMRIGWLITDCEYVTHQWRVAGSLGVSADLLGVR  
 DMQYNGFGPAEQALVLAATDDVVREGVSAQSWACERELHCDKVVLLIELVTVISAR  
 MVASILHSLEVPEDGVSSWPPDGLSPR"  
 misc\_feature

```

gene
complement(1859..2293)
/locus_tag="Rv1532c"
CDS
complement(1859..2293)
/locus_tag="Rv1532c"
/function="UNKNOWN"
/notes="Rv1532c, (MTCY07A7A.01c), len: 144 aa. Conserved
hypothetical protein, similar to P20378|YPHR HALHA
Hypothetical 15.6 kDa protein from Halobacterium halobium
(151 aa), FASTA scores: opt:152, E():4.5e-05, (30.1%
identity in 103 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAB11594.1"
/db_xref="GI:2370321"
/translation="MSDPLTAQEQHKRQAVRELMPTPFPGGLGIVFERYEPDDVVI
RLPFTDLTNDGTYFHGGVIVASVMDTAGAAANSHDFDRGTRAAVTAMSIOYTGAAK
RCDLLCHARTARRRKELTFTEITATDPDGNVAHAVQTVIV"
2353..3480
/locus_tag="Rv1533"
2353..3480
/locus_tag="Rv1533"
/function="UNKNOWN"
/notes="Rv1533, (MTCY07A7A.02), len: 375 aa. Conserved
hypothetical protein. Similar to 2NPD NEUCR|Q01284
2-nitropropane dioxygenase precursor (378 aa), fasta
scores: opt: 279, E(): 9.1e-11, (31.3% identity in 256 aa
overlap). Also similar to Mycobacterium tuberculosis
hypothetical proteins Rv1894c, Rv0021c, Rv3553, Rv2781c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAB11595.1"
/db_xref="GI:2370322"
/translation="WRTKRVALLGAEPFICAFSHCRDVAASVNGGPGILGAHVAHSP
KRLEETWIEHTGKPYGVLDLIPKYGABQGGIDAQOARELIPGHRFTFVDDLL
VRYGIPATVDRQSSAGSLHISPKGYQLDVAFHDIRLIASALGPPDPLVERAH
NHDVLVAALCTAQRHRAAGVDLI VAQGTGAGHTGEVATVVLVPEVVDVATPPT
VLAGGIARGQIAAALGAGVGCWSVLTTEATPPVVKDFLAATSSDVTVRSR
SLTGKPAKMLKATMDEWDRESDPDPLGNPLQSAIVSDPQIRINQAQGPQAKARELA
TYFGVQVYSLDVRVSRAVLDVMEEFITDVQLQGLVQR"
3477..4154
/locus_tag="Rv1534"
3477..4154
/locus_tag="Rv1534"
/function="Possibly involved in a transcriptional
mechanism"
/notes="Rv1534, (MTCY07A7A.03), len: 225 aa. Probable
transcriptional regulator, similar to YCDC ECOLI|P75899
hypothetical transcriptional regulator from Escherichia
coli (212 aa), FASTA scores: opt: 166, E(): 9.8e-05,
(24.2% identity in 219 aa overlap). Contains PS01081
Bacterial regulatory proteins, tetR family signature and
helix turn helix motif (aa 41-62)."
/codon_start=1
/transl_table=11
/product="Probable transcriptional regulator"
/protein_id="CAB11596.1"
/db_xref="GI:2370323"
/translation="MSRASRRRRVSDSKSQRRDEILAAKIVFAHKGFGHATTVAD
IAKQAGLAVGLTYWYFDKDLDFHALMAGEEELAAHVAELARVGGSTEAPLRALLQ
RAVQATFEFFEDKATVLLFDVAIGRFEELGIVVERFIDIEAVVVAQRGE
VVEASPMNAAYTLAALVGQLAHRLNTDDNTVTAQVADPVVSLVDGLRPRALAVGAR
GGRAART"
3582..3674
/locus_tag="Rv1534"
/notes="PS01081 Bacterial regulatory proteins, tetR family
signature"
4719..4955
/locus_tag="Rv1535"
4719..4955
/locus_tag="Rv1535"
/function="UNKNOWN"

misc_feature
3582..3674
/locus_tag="Rv1534"
/notes="PS01081 Bacterial regulatory proteins, tetR family
signature"
4719..4955
/locus_tag="Rv1535"
4719..4955
/locus_tag="Rv1535"
/function="UNKNOWN"

gene
/notes="Rv1535, (MTCY07A7A.04), len: 78 aa. Hypothetical
unknown protein."
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAB11597.1"
/db_xref="GI:2370324"
/translations="MTAAALHNDVVTVASAPKLRVVRDVPAPASKKVARRLDAQPFGT
GGDPLVDGAARLLSIPLRHLIYAALWRVGLLEVQA"
5262..8387
/genes="iles"
/locus_tag="Rv1536"
5262..8387
/genes="iles"
/locus_tag="Rv1536"
/EC_number="6.1.1.5"
/functions="charging ile tRNA [CATALYTIC ACTIVITY : ATP +
L-ISOLEUCINE + TRNA(ILE) = AMP + DIPHOSPHATE +
L-ISOLEUCYL-TRNA(ILE)]."

Query Match 4.6%; Score 23; DB 1; Length 347496;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 ACGCTGATCTTCGACTACCCAC 215
|||||
Db 150140 ACGCTGATCTTCGACTACCCAC 150162

RESULT 20
AK063543
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:001-117-C12, full
insert sequence.
DEFINITION AK063543 723 bp mRNA linear PLN 24-JUL-2003
ACCESSION AK063543.1 GI:32973561
VERSION F1I CDNA; oligo-capping.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-Length cDNA Project Team.,
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yanagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shihiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ootomo,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yamashita,K., Riken.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K.,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Itoh,M.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
Sato,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE 2 (bases 1 to 723)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
Kanagawa,S., Kawai,H., Kawagashira,N., Kawai,J., Kawamata,M.,
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,

```



```

Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Inamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, K., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES             source
1. 1019
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="mRNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /clone="002-125-E05"

ORIGIN
111168 CGGTGCTCGGCCACAGCGGCGC 568

Query Match          4.4%; Score 22; DB 8; Length 1019;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CGGTGCTCGGCCACAGCGGCGC 83
|||||
Db 589 CGGTGCTCGGCCACAGCGGCGC 568

RESULT 22
LOCUS AC135598 137312 bp DNA linear HTG 16-OCT-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 3 clone
OSJNB0021K20, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC135598
VERSION HTG: HTGS PHASE2; HTGS ACTIVEFIN.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 137312)
AUTHORS Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
Gausberger, K., Burgess, S., Jarrahi, B., Shvartsbeyn, M., Brenner, M.,
Ciecko, A., Pal, G., Vanaken, S., Hansen, C., Utterbach, T.,
Feildlyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
Salzberg, S. and Fraser, C.
TITLE Oryza sativa ssp. japonica cv. Nipponbare OSJNB0021K20 BAC genomic
sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 137312)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Oct 16, 2003 this sequence version replaced gi:24418097.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 68692: contig of 68692 bp in length

```

```

* 68693 68792: gap of unknown length
* 68793 137312: contig of 68520 bp in length.

FEATURES             source
1. 137312
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="genomic DNA"
   /cultivar="Nipponbare"
   /sub_species="japonica"
   /db_xref="taxon:39947"
   /chromosome="3"
   /clone="OSJNB0021K20"

ORIGIN
Query Match          4.4%; Score 22; DB 2; Length 137312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CGGTGCTCGGCCACAGCGGCGC 83
|||||
Db 111168 CGGTGCTCGGCCACAGCGGCGC 111189

RESULT 23
LOCUS BX572597 349737 bp DNA linear BCT 18-DEC-2003
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 5/16.
ACCESSION BX572597 BX571963
VERSION BX572597.1 GI:39648199
KEYWORDS complete genomes.
SOURCE Rhodopseudomonas palustris CGA009
ORGANISM Rhodopseudomonas palustris CGA009
Rhodopseudomonas palustris CGA009
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiales; Rhodopseudomonas.
REFERENCE 1 (bases 1 to 349737)
AUTHORS Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
Land, M.L., Pelletier, D.A., Beatty, T.J., Lang, A.S., Tabita, F.R.,
Gibson, J.L., Hanson, T.E., Torres y Torres, J., Peres, C.,
Harrison, F.H., Gibson, J. and Harwood, C.S.
TITLE Complete genome sequence of the metabolically versatile
photosynthetic bacterium Rhodopseudomonas palustris
Unpublished
REFERENCE 2 (bases 1 to 349737)
AUTHORS Larimer, F.W. and Harwood, C.S.
CONSRTM Rhodopseudomonas genome consortium
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
genome consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov

FEATURES             source
1. 349737
   /organism="Rhodopseudomonas palustris CGA009"
   /mol_type="genomic DNA"
   /strains="CGA009"
   /db_xref="taxon:258594"
   complement(232..1425)
   /gene="carA"
   /locus_tag="RPA1276"
   complement(232..1425)
   /gene="carA"
   /locus_tag="RPA1276"
   /EC_number="6.3.5"
   /function="InterPro
IPR000991:IPR001317:IPR001687:IPR002474:IPR006274 COGS
COG0505"
   /note="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
   /codon_start=1
   /evidence=not experimental
   /transl_table=11

```

```

/product="carbamoyl-phosphate synthase small subunit"
/protein_id="CAE26719.1"
/db_xref="GI:39648200"
/translation="MTNSASHPAWPDHKKHTALLIADGVTFVFGGLGARGHAVGEVCF
NTMTGYSIELTDPSTAGOLITFTPHIGNVGTNDEIETVNNAAITPGARGVILRSAL
TDPNYSRSHLDLWKARIGIIGLSGIDTRALTALIRSGMPNVAIAHSPTGEFDLHA
LKEAREWPLGEMDLVPMTVTAQRTWDETPTWAGESFGRODKPEFNVAIDYGTGR
NIRLLAGGCKVTVVPATTSADILAMKPDGVFLNSGPDPAATCKYAVPVHGVIE
TGVPFGICLGHQMLGGLGKTVKMHQGHGANHPVKDLTKGKVEITSMHGEAVDK
STLPANVQOHTHLSLFDNSCGIALADKPVFSVOYHPFASPGQDSHYLFRFTDLNRA
KKAAS"
gene 1574. .2278
/locus_tag="RPA1277"
CDS 1574. .2278
/locus_tag="RPA1277"
/function="InterPro
IPR000073:IPR000379:IPR001031:IPR003089 COGs COG0596"
/notes="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible beta-ketoadipate enol-lactone
hydrolase"
/protein_id="CAE26720.1"
/db_xref="GI:39648201"
/translation="MTSLPVLVPLGAGSPRIYAFIFPALWACGSPVTVANHIRDS
MAAIRILAEAPRFALAGSHGGYIAFEIMQAPDRVAKALINTQALPDSPEATE
RRTOIGEAERGLHVDLPEYGFVHPSRDDAVLRTVVHEMGDDVGAAGFVQORA
IARTDSRHTLGTIGTFLIVISGDTTTPISLSQEMADGITGAKLLIIPDCGHLPOI
EOPAAVAALADWLNR"
gene 2436. .2891
/locus_tag="RPA1278"
CDS 2436. .2891
/locus_tag="RPA1278"
/function="InterPro IPR003789 COGs COG1610"
/notes="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="GatB/Ygey"
/protein_id="CAE26721.1"
/db_xref="GI:39648202"
/translation="MURDDINNAVKAMKAKDERKLSLRMNSTINKADIARGQGK
PLSDGDLISLQKMIKQKQESVELYDKGRAELADQRAETIAVIQAYLPQQMSDDEV
KAAIAATISSETGAAGIKDMGKVTGALKAKYAGQMDFGKAGMVKAAALTG"
gene complement(2979. .3287)
CDS /locus_tag="RPA1279"
complement(2979. .3287)
/locus_tag="RPA1279"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26722.1"
/db_xref="GI:39648203"
/translation="MDAFAPQHUIHVVSAGWOAPLHTVVNGCAGAPRSLVGLSRL
RRAPDRRGHDPAAGLPQORQAGDGLSLISHSGFASDRDPGLIQFSRRAQORGARRP
"
gene 3478. .4326
/locus_tag="RPA1280"
CDS 3478. .4326
/locus_tag="RPA1280"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26723.1"
/db_xref="GI:39648204"
/translation="MSISGKVDPVWRPTITIGDIAELGQLGRDFOQAPLYGLAFAPY

```

```

AAGGLILACLTAFAHMYLALPLGAGFALIGPFVALGLYEVSQRQREAGKPSLIQIVG
LMRSSELGMAFVTLFLVIMYOVRLLIALFLGVGASFGSLQEFISAVLTNEGIV
FIAGNCVGACIALVSLTAVSFPLLRDQDVFVTAMTSVRAVVKSPLEPMIGWAAT
IVLILASLPFLGLVVLTPVLGCHATWHLVKRIVAPVAELPDTADTEASNNVAMP
KKAATG"
misc_feature order(3580. .3648,3658. .3726,3805. .3858,3868. .3936,
3955. .4023,4132. .4200)
/locus_tag="RPA1280"
/notes="6 probable transmembrane helices predicted
byTMHMM2.0"
gene complement(4440. .4913)
/locus_tag="RPA1281"
CDS complement(4440. .4913)
/locus_tag="RPA1281"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26724.1"
/db_xref="GI:39648205"
/translation="MSKPEFAVTVYTRTPDLRHLALTDADETRYVMDCITLRSDDTV
GSEMTMERGEIKNRCTIYESDPHRLAYDWVSWDPAMQEKPSVTVYIEPQGLV
KUTVTHQNTQSGTTLPSISFGWVVLASLSILETGQLPFPAPQSHAAESAHV"
gene complement(4906. .5220)
CDS /locus_tag="RPA1282"
complement(4906. .5220)
/locus_tag="RPA1282"
/function="InterPro IPR001845 COGs COG0640"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible transcriptional regulator, ArsR family"
/protein_id="CAE26725.1"
/db_xref="GI:39648206"
/translation="MDEVKALADGSRRLDLRHARGQTLSDLCGLAMTQAVTK
HLAILEAANLVATVKHREKLYLNPMPHEIGERWIRKFERGLAALSDLKRLKETR
DE"
gene 5349. .5966
/locus_tag="RPA1283"
CDS 5349. .5966
/locus_tag="RPA1283"
/function="InterPro IPR001123 COGs COG1280"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="homoserine/homoserine lactone/threonine efflux
protein"
/protein_id="CAE26726.1"
/db_xref="GI:39648207"
/translation="MSLEMYLTFVTACIVIALIPGPVTVIANGRHGTRAAALNTIA
GAQLGIGLIALVAAGLTTLMATMGVFDWVFLGAAYLVLGILIRREVEGLSTDQ
PPTPRGPGFLQGFVLLANPKVLFVFGAFIPQFVDMQRETVPQVALLGVTFVIAAM
TDAYALLAGRAKTMFSQORILLRISGISGFMGGIWLALSRAK"
sig_peptide 5349. .5426
/locus_tag="RPA1283"
/notes="Signal predicted by SignalP 2.0 HMM (Signal
peptideprobability 0.832) with cleavage site probability
0.732 atresidue 26"
misc_feature order(5367. .5435,5463. .5531,5568. .5621,5679. .5747,
5784. .5852,5895. .5954)
/locus_tag="RPA1283"
/notes="6 probable transmembrane helices predicted
byTMHMM2.0"
gene complement(6090. .6887)
/locus_tag="RPA1284"
CDS complement(6090. .6887)
/locus_tag="RPA1284"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved unknown protein"
/protein_id="CAE26727.1"
/db_xref="GI:39648208"

```

/translation="MDGVRLNFLSLDAGEPPVIERNONGOSPELLVCDHYGROI PAQL  
GDGLPRAETRIHAWDGTAGVAERMSDALGAHLIAQYSRIVIDCNRPSPAESIP  
LVSEATIPNDGISEHAEERRLIFDPYHOREIAIEDROLSGOPTVLLAHSHSTS  
VYHGMARPHIGTLYQCTVIPPULLAELREPELVIGNQPIAVSDGDTYIPVHGT  
ARGLITSGBIRQDITDEAGQVEWASRLTRVLSTIADTLRQQOALR"

Query Match 4.4%; Score 22; DB 1; Length 349737;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 GCTGCTCGGCTCGCCGACACC 378

Db 135785 GCTGCTCGGCTCGCCGACACC 135764

RESULT 24  
AFL144053 727 bp DNA linear BCT 12-JUL-1999  
LOCUS  
DEFINITION Micromonospora inyoensis macrolide type polyketide synthase genes,  
partial cds.  
ACCESSION AFL144053  
VERSION AFL144053.1 GI:5442290  
KEYWORDS Micromonospora inyoensis  
ORGANISM Micromonospora inyoensis  
SOURCE Bacteria; Actinobacteridae; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1 (bases 1 to 727)  
AUTHORS Hyun, C.G. and Suh, J.W.  
TITLE The use of PCR to isolate new macrolide type polyketides from  
actinomycetes

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 727)  
AUTHORS Hyun, C.G. and Suh, J.W.  
TITLE Direct Submission  
JOURNAL Submitted (19-APR-1999) Department of Biological Science, Institute  
of Bioscience and Biotechnology, Namdong San 38-2, Yongin,  
Kyunggi-Do, Korea

FEATURES  
source Location/Qualifiers  
1..727  
/organism="Micromonospora inyoensis"  
/mol\_type="genomic DNA"  
/strain="ATCC27741"  
/db\_xref="ATCC:27741"  
/db\_xref="taxon:47866"  
<1..567  
/note="type I PKS"  
/codon\_start=1  
/transl\_table=1  
/product="macrolide type polyketide synthase"  
/protein\_id="AAD43313.1"  
/db\_xref="GI:5442291"

/translation="ASRDLGDFSVLSLELRNLAATGIRLASGLLFNHPPTAALAGH  
LRHLVGGVDPAAATRADADAEPIAVICPGDVRSSEDLRLVAGGVDAIGEL  
PTNRGWDLEIGDFDPGPGVKGSRHGGFLHRAEDFEFFGISPREAAMDPOQRLL  
LETSEWAEFAGIDPTACAAAPACSWG"  
567..>727  
/note="type I PKS"  
/codon\_start=1  
/transl\_table=1  
/product="macrolide type polyketide synthase"  
/protein\_id="AAD43314.1"  
/db\_xref="GI:5442292"  
/translation="MSQYGPRLYTSAGSDGYRLTGSTASVASGRISYVFGFEPAP  
TVDTACSSSL"

QY 165 CGCGGCCACCGGCTCGCGCT 185  
|||||  
Query Match 4.2%; Score 21; DB 1; Length 727;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 CGCGGCCACCGGCTCGCGCT 80

RESULT 25  
SGR300302/c

LOCUS  
DEFINITION Streptomyces griseus partial ORF1, canA gene, canC gene, canF gene,  
canT gene, canRA gene and canRB gene.

ACCESSION AJ300302

VERSION AJ300302.1 GI:12231153

KEYWORDS ABC-transporter; aminotransferase; can RB gene; canA gene; canC  
gene; canF gene; canRA gene; canT gene; Cho-like protein;  
cytochrome P-450; ferredoxine; PABA synthase; thioesterase.

SOURCE Streptomyces griseus

ORGANISM Streptomyces griseus

Bacteria; Actinobacteridae; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1

AUTHORS Campelo, A.B.  
JOURNAL Thesis (2000) Department of Microbiologia, Universidad de Leon,  
Leon, Spain

REFERENCE 2

AUTHORS Campelo, A.B. and Gil, J.A.  
TITLE The candidin gene cluster from Streptomyces griseus IMRU 3570  
JOURNAL Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)  
MEDLINE 21642576  
PUBMED 11782498

REFERENCE 3 (bases 1 to 39314)

AUTHORS Gil, J.A.

TITLE Direct Submission  
JOURNAL Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de  
Leon, Campus de Vegazana s/n, 24071, SPAIN  
FEATURES  
source Location/Qualifiers  
1..39314  
/organism="Streptomyces griseus"  
/mol\_type="genomic DNA"  
/strain="IMRU 3570"  
/db\_xref="taxon:1911"  
complement (<1..1785)  
/note="ORF1"  
/codon\_start=1  
/transl\_table=1  
/product="Cho-like protein"  
/protein\_id="CAC22112.1"  
/db\_xref="GI:12231154"  
/db\_xref="SPTREMBL:O9EWC6"  
/translation="MLLRTETIALVDALGAATEGRSSIVLLTGLIGRSALLORL  
SCLADDDVVRVRAHRAAMEQDFAGVVRQLFETLLGDSPEADRDMDASPARH  
VLADDAAPPDAQIAATEAVLHGLLSLVANSADSRLLILVDDLDQSDVFSRLWLY  
LAKELHGLRAVVVCAIRDGDPDRSHHTLVREIREAATQTLRPAISLSTHELVRHFG  
EAGDDEFVQACHASVGNPLFLSLVGLTGLRRPLAEHAEATRRPLRPSQLRRELAS  
ILRTQAPVRDLAAAIILGQSDAPTLRLAGLDSIGYAGALRALGALGALAAAPDEP  
RFHRSVRDAESTLTWQREHDEAALLYAAGSPAEQVAQOLMAVVRQPPAVD  
VIRAAATLARGAPDTNAGLYCRALDSDSPAGVGRGLLVELGTAEFGFPLACERH  
IAQATLLPEPRDAVALRLSPALGPAPLAVDLLKQAEDLGPAGLGTGTDRLA  
LRLEARLHCHEDPHELAASAHRLKEIGPEPPVDTGGERELTALLGAATLGCHLPA  
SEVVRAGRVLDREPATSARHTSLPLAI"

9525..10583  
/gene="canA"  
9525..10583  
/gene="canA"  
/codon\_start=1  
/transl\_table=1  
/product="aminotransferase"  
/protein\_id="CAC22113.1"  
/db\_xref="GI:12231155"  
/db\_xref="GOA:O9EWC5"  
/db\_xref="SPTREMBL:O9EWC5"  
/translation="MAFTHPVSRPALDGRLELYVSDAVSGGWSISQGYVRRRFEBAFA  
EWNGVAGVACSSGTALTLRALNIGDPGEIVPEFTMTWASAWAVTGTATPVFVD  
CGDLDINDVITIEKLIPTRAVMPVHVYGRCDMDAVMDLALQYNLRVSDSAHAG  
VRPVGDIACSLFANKIITAGGGVCLTDFRLAEQLARLMAFTRDHSFLHKLAY  
NYRMTAMQGAVALAQTRELDEILATREIARYDAGLKDLPGLTLPMPARDVLNMYDLR

```

AERREURAHLDIERGIERLFFKPMRQPGYLDVPMVPTLNAHRFSEBGLYLPTHTGLT
AAQDEYITGAVRDFYRAG"
10678. .11859
/ gene="canC"
10678. .11859
/ gene="canC"
/ codon_start=1
/ transl_table=11
/ product="cytochrome P-450"
/ protein_id="CAC22114.1"
/ db_xref="GI:12231156"
/ db_xref="GOA:Q9EWC4"
/ db_xref="SPTREMBL:Q9EWC4"
/ translation="MTSPGTVVDFFPRTREPLPLSQAEHRKQNGLVOTHLPNGR
PIWLTHREDVRAVLTHPRISANPDNEGPNVGTMGVQKQEQIIPGFVGLDSPEHR
FRKVLIEFTVRRVRELRPAIERTVDERIDAMLAGNADLVNDFALPVPVSLVISALL
GVPSADRFPSRTVAIRSTDEERAEATRQLRYINRLIIVIKHREGEDLISRL
LSTGKLDEELSGVLLLIAGHTETANNIGLVVTLISHREWIGDDRLVELLRLHS
VADMVALRVAVDDVEIAGQTTKRGEGIVPLIASANHDTEAFGCPHAENPETERRHA
FGYGVHQLGQNLVRVMEIAYKLFERIPELRLAVPEDQLAYKXDGLFGLHELPIYR
W"
11906. .12100
/ gene="canF"
11906. .12100
/ gene="canF"
/ codon_start=1
/ transl_table=11
/ product="ferredoxine"
/ protein_id="CAC22115.1"
/ db_xref="GI:12231157"
/ db_xref="GOA:Q9EWC3"
/ db_xref="SPTREMBL:Q9EWC3"
/ translation="MRVTVDSQCVGAGQCQVNAPEVFDQDDGGVVLLRAEPDERDH
EAVTAGDLCPASVYLOED"
12223. .12993
/ gene="canT"
12223. .12993
/ gene="canT"
/ codon_start=1
/ transl_table=11
/ product="thioesterase"
/ protein_id="CAC22116.1"
/ db_xref="GI:12231158"
/ db_xref="GOA:Q9EWC2"
/ db_xref="SPTREMBL:Q9EWC2"
/ translation="MTTADDTAARLWRYHPAEADAIVRLVCFPHAGGSASFVHPVSR
FAPGAEVYSQYQGRQKRPFCVPLDGLTLADLITEQLPLDERPTVFHSMGAALA
FETAWRLQKAGPRTVIASGRGSPSTTAEKVHTRDDGIVAEKRLNGTAAGVLGD
BEILMVALPALRGDVRATETVTCPPDRRLACGLVLTGDDDDPLTTVEAEWRDHTTG
PRLRVFGGHPFLTQHLDAVNTAIAQALHPPRAAPA"
13190. .15361
/ gene="pabAB"
13190. .15361
/ gene="pabAB"
/ codon_start=1
/ transl_table=11
/ product="PABA synthase"
/ protein_id="CAC22117.1"
/ db_xref="GI:12231159"
/ db_xref="GOA:Q9EWC1"
/ db_xref="SPTREMBL:Q9EWC1"
/ translation="MRTLLVDNYDSFTYNLPHYLFRANGREPEVINDDPAWRGLLD
AFDNLVSPGTPHRRPFDGFCARIAEGRULPVLGVLGHQGMALHAGRVGRAPPE
RGRGTSAVRHDTGLFEGLPQVLEVRVHSLATLPELPELEATWSDEGVMALRHRT
LFLGWGVPESIGTQDGHRLANFRDLTERGTRTHGGAGHGTLPPAPARETKAT
TCTPRLRVIAKSLPDAEVAFDSLPRTGDHPFLWDSRRPGELGOLSMMDGASGP
LARTAKDVHACTVTVRADGASTVESAFLTWLENDLACLTEVEPLPAFALGWVC
LGYELKACDGGAAHRSPPDAVLPADRALVLDHRTTYLLALVEDDAEAEARWL
AAASATLAVAGEPEPCPEAPVCTTGPVELRHRDGLYKLDVCCQEQEIAAGETVC
LTNMAEDTLFPFAEALRVSRPAPFAFLDFPGPMVLSSSPERFLRDRHGRMS
FIKGTFRPGATPQDAADLVRALATCEKRAENLMIVLVRHDLGRCAEVSADVDPV
FOVETVATVQVSTVARTLRDSSPVAAVRAAFPGGSMTGAPKRTMOIIDLGGGP
RGVYSAGIGYFSLTGAVDLSIVIRTVVLSGRLRYGVGGAVIALSDPDDEFETAVKA
```

```

APLLRLILDITAPPGREAPFGKDLIDGPDGDTAGAPKDLVLPQ"
20829. .21836
/ gene="canRA"
20829. .21836
/ gene="canRA"
/ codon_start=1
/ transl_table=11
/ product="ABC-transporter"
/ protein_id="CAC22118.1"
/ db_xref="GI:12231160"
/ db_xref="GOA:Q9EWC0"
/ db_xref="SPTREMBL:Q9EWC0"
/ translation="MNGHLPPALPAPGVSAAAGSGDLMIEAGVSKAYGTVHALDSV
SLAVPRCTVLGLGHGAGKTTIVDILTTALPTTSGRARVAGYDVADKPIRRRIQL
TGQASVDQAQLSGYDNLVLIARLAGARRARVARELLELFLTEVADRASSYSGG
LRRDLDAVLSVGAPEVFLDEPTTGLDPPSRINLWEIVEGLVQGTVLLTTQYLEE
ADRLADRIANLSAGRVAAAGTAPBLKATVGRRTVTLTLETGEDVSAARTALRGAFAP
VDGEARTVTVVIDATREIADIIRSLDQAGVEASELNFCEPTLDDVYLLTAEQAAGRTR
A"
21970. .22776
/ gene="canRB"
21970. .22776
/ gene="canRB"
/ codon_start=1
/ transl_table=11
/ product="ABC-transporter"
/ protein_id="CAC22119.1"
/ db_xref="GI:12231161"
/ db_xref="GOA:Q9EWB9"
/ db_xref="SPTREMBL:Q9EWB9"
/ translation="MPRAVAPAFGGSLSFTQIRVLTRGSLRAMVTDPGIVLFGLIQL
VVILFVLTVQFSKMGMPHPFDGVSYLDPVLPAVDNAOASOMSGVGLVEDOKNGI
VARLSPVPHGALLAARSLNGLVRSAYQAVAMALANTVLGYSNGGAELASGL
TLFISFSGWAFIAGAWIRRAEPLQNALIVIFPFASASAYVPAADLPWLGAVAS
VNPLTYAIDATRALDVPGLDHALPAVVICAVIAVGVGMIAAVRGFRRLP"
Query Match 4.2%; Score 21; DB 1; Length 39314;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 121 GGCTTCGACTCGCTGGCCGCG 141
|||||
Db 34672 GGCTTCGACTCGCTGGCCGCG 34652
|||||
RESULT 26
AF263245 AF263245 47981 bp DNA linear BCT 17-SEP-2000
LOCUS Micromonospora megalomicea subsp. nigra megalomicin biosynthetic
DEFINITION gene cluster, partial sequence.
ACCESSION AF263245 GI:10179840
VERSION AF263245.1
KEYWORDS Micromonospora megalomicea subsp. nigra
SOURCE Micromonospora megalomicea subsp. nigra
ORGANISM Micromonospora megalomicea subsp. nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1 (bases 1 to 47981)
AUTHORS Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.
TITLE Biosynthesis of the anti-parasitic agent megalomicin:
transformation of erythromycin to megalomicin in Saccharopolyspora
erythraea
JOURNAL Mol. Microbiol. 37 (4), 752-762 (2000)
MEDLINE 20430101
PUBMED 10972798
REFERENCE 2 (bases 1 to 47981)
AUTHORS McDaniel, R. and Volchegursky, Y.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
Location/Qualifiers
1. .47981
source
```

```
/organism="Micromonospora megalomicea subsp. nigra"
/mol_type="genomic DNA"
/strain="NRRL3275"
/sub_species="nigra"
/db_xref="taxon:136926"
/complement (<1. .144)
/gene="megI"
/complement (<1. .144)
/gene="megT"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyglucose-2,3-dehydratase"
/protein_id="AAG13906.1"
/db_xref="GI:10179841"
/translation="MGDRVNGHATPESTQATRFTRHGPPPTATDDVHDWLAHRAAE
HRLK"
928. .2061
/gene="megDVI"
928. .2061
/gene="megDVI"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"
/protein_id="AAG13907.1"
/db_xref="GI:10179842"
/translation="MAVGDRRLRGRLGELMARGLYWFGANGDLYSMLSGRDDDPWTW
YERLRAAGRPVAGTAVGDRHTAAEVLADPGFTGPPDAARWQVAHCPAAWPA
GPREFVATEDAASVTVDADLQQRCAVLTELGRFDLVNDFAREVPVIALGTAPA
LKGVDPRLRSWTSATRVCLDAQVSQQOLAVTEQALTDALDEIDAVTGGDRAVLGVV
AELANVTGNVLAVTELPALADDPETATRVTEVRSFTSGVHLERRTHAASRR
VGVDPVTGGEVTVVAAANRDPEFTDPFDVDRGDAELISLSPSPRTDLDALV
ATLATAALRAAAVPLPRLSRSGPVIRRRSPVARGLSRCPEVL"
2072. .3382
/gene="megDI"
2072. .3382
/gene="megDI"
/codon_start=1
/transl_table=11
/product="TDP-megosamine glycosyltransferase"
/protein_id="AAG13908.1"
/db_xref="GI:10179843"
/translation="MRVFFSSMAVNSHLFGLVFLASAFQAAGHEVRVVASPALTDVTT
GAGLTVPGDDVDELVEHAGAGQDIVEYMRITLDWDQSHSTMSWDDLLGMOTFTPT
FFALMSPSLIDNGVFCRWRPEDIWPEPLTAAPIAAVGTGTGHARMLWGPDAVTR
AROSFLRLAQVEHREDPLAEWDMTLRRFGDHLSEELVLGQWTVDPILPPL
RIITGVTVGWRVYVNGPSPVAVALLRPERRVCITLGSRSRHGIGOVISGEMLD
AIADIAEFVATDDQLGVGVGPANVTAGFVPMNVLPTCAATVHHGCTGSWLTA
AIGHVFOILISDADTEVHAQQLQDLGAGLSLPVAGMTAEHLRGAIERKVLDEPAYRLGA
ERMRDGMRTDPSPAQVVGICQDLAADRRAARGRPRTAEPLPR"
3462. .4634
/gene="megY"
3462. .4634
/gene="megY"
/codon_start=1
/transl_table=11
/product="mycarose O-acyltransferase"
/protein_id="AAG13909.1"
/db_xref="GI:10179844"
/translation="MWTSTNLDYTABALNSLTGMRTVAAFLVFFTHVLSRLPNSTVY
YADGLDAFWQITGRGVSVFFLLSGFVLTSARASDSVWGFRRRVCKLFPNHLVTAF
AAVFLVTVQAVSGEALINLLLIHAWFFALEISFGINPVWSLACEAFYLCFLPLF
LFWTGRRLPRLMNAWAAVFAAIWAVPVADLLPSPPLTPGLEYSALQDMFLYTFP
LRLSELTILGILALITGRWINVGLLPVLLPVPFVASFVLPVGVVATISSMSLTA
LVLLIASCATADLOOKTEMRNVWMLGDVSPFALYVWHTLVIVYGADLLGFSQTEA
PLGLALFWIPLFLAVSLVLSWLLYRFVELPVRNWARPASRRKATEPEQTSSRR"
4651. .5775
/gene="megDII"
4651. .5775
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG13910.1"
```

```
/db_xref="GI:10179845"
/translation="MTTYVWSYLLBYERERADILDVQKVFASGLILQOSVENFETE
VARYGHIAHCVGMDGTNAVLALSLSVGVGRDDEVTVVTSNTAAFTVLAIDEICARPVF
VDVRDEDYLMDDTDLVEAAVTPRTKALVPVHLXQCVDMTALRELARGLKLVEDCAQ
AHGARDRELAGTMSDAAAFSPYFKVLGAYGGGAVTNDDEETARAYYGMEE
VYIVTIRPGHNSRLDEVQAEILRKLRLDLDIVAGRAVAVQRYVDQGLADLQDSHGLEL
PVVTDGNEHVFYVYVHRPRDEIILKRLDGDYDILNITSYPMFPVHTMTGFAHLGVASG
SLPVTTERLAGEIFSLFMYPSLPHLDQRLQVIEAREVITGL"
5822. .6595
/gene="megDIII"
5822. .6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N'-dimethyltransferase"
/protein_id="AAG13912.1"
/db_xref="GI:10179847"
/translation="MPSHSTSSSTDVAIYERADIIYHDFYHGRGKGYARAEADALVEVA
RKHTPQATLLDVACTGSHLVELSDVSFEVGVGVDLSAAMLAATAANDPGREVRHQDM
RDFSLDRPDDVVTCMFSYGLVDEAEELDRAVANLAGHLAPGGLTVPEWPFETPRP
GWGADLVTSGRRI:SRMSHTVPAGLPDRTASRMTIHYTVGSPGAGIEHTEVHVMTL
EAAAAYEQAFQAGLSYVGHDLFSPGLFVGVAABEGR"
6592. .7197
/gene="megDIV"
6592. .7197
/gene="megDIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="AAG13911.1"
/db_xref="GI:10179846"
/translation="MRVEELGEGVFTFTQTFADERGVTGTAQEDVFAALGRPLF
PVAQSTTRSRRGVGVGVTFTMPGSMKYVYVARGAMDFAVDIRPGSPTFGRAEPF
ELSAESMGLYLPVGMGHLFVLEDDTTLVILMSAGYVDPKERAHPDLDELALPIPA
DIDLWMSERDRVAPTLREARDQGLPDIYACRAAAHVVRT"
7220. .8206
/gene="megDV"
7220. .8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translation="MVLGASGLGSATVHALDLFVRVRLVARREVVVPSGAVADYE
THRVDTLTPCALAEVAVADAVFPFAAQIRGTSQWRISEDDVVAERTNVLGVADLIAY
LSRSHAPVVVPGSNTQVGRVTAGVIGSEODHPGVYDROKHTGEQLKEATRAAG
AIRATSLRPLPVGVPAAGTADRGVVSIMIRRALTQGPLTMHMDGTVRRELLVYTD
ARAFVTDHADALAGRHELLGTGRSWPLGVEVQAVRSVRHGTGEDFVSVVVPFPA
HMDPSDLRSVEDPAPRTAVTGMRTVTWAEAVDRTVAALAPRAAAPSFS"
complement (8228. .9220)
/gene="megDVII"
complement (8228. .9220)
/gene="megDVII"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AAG13914.1"
/db_xref="GI:10179849"
/translation="MGITGASRVVRGSRALHTSRILWLTGTVNFSGRVTDALRLMD
HALRGVNCIDADIYWRLYKGTBELVGRWFAGGGRREETVLAATKVGSEMSRVN
DGSLARHIVACENSIRLGVGHDIIYQTHHIDRAAPDEWVQAAHLYGSGKGVY
GSSLNAGWHLIAAQESAARNLLGMSHQCLYNLVRHPBELDLPAAQAYGVGVFWS
PLHGGLLSGLVLEKLAAGTAVKSAQRAOVLPAVPLVEAYEDYCRRLGADPAEVGLA
WVLSRPGILGAVIGPTPEQLDLSALRAAELTLGEEELRELEAIFPAVADGPVE"
complement (9226. .10479)
/gene="megBV"
complement (9226. .10479)
/gene="megBV"
/codon_start=1
/transl_table=11
/product="TDP-mycarose glycosyltransferase"
/protein_id="AAG13915.1"
```



```
/db_xref="GI:10179850"
/translat="MRVLLTSFAURTHFOGLVPLAWALHTAGHDVRVASOPELTDVVV
GAGTSLPLSGDHLRFIDSPAAQVHYRTTDLDFARRGPELRSMWFLHGIIEATSRF
VFPVNDVSFVDELVEFAMWRPDLVLEPPTFAGAVAAKACGAARHLLWGLSTGY
FRSQDLRGQRADDPDPLUGGWLTEVAGRFGLDSEDLAVGOWSDQLPESRLET
GLESHVTRTLDPYNGSSVVPQMLRSDGVRVFCFTGGYSALGITSNPQEFURLTATLR
FDGIVTVTRSLGDPASVDPNVLVDVFPVNNILLPGCAAVIHGGAGSWATLHHGVQ
ISVAHEMDVLVRGORTAELGAGVLRDPEDVDATLWQALATVVEDRSHAEAKLRQE
ALAAETAEVYVLEALAHQHRADR"
complement (10483..11424)
/genes="megBIV"
complement (10483..11424)
/genes="megBIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13916.1"
/db_xref="GI:10179851"
/translat="MTRHVTLLGVSGFVGSALLREFTHPLRLRAVARTGSRDQPPGS
AGIEHLRVLDLEPGKRVQAVDQVTVHLVAYAGSGTWSRAATVPAERDVNAGIMRD
LVAALRABGPAPVLLPASTTQAAANPAAPSAOAKIEAERILRQATGVDGVILR
LPAIYHSGSPSGQGRGVVTHAMIRALAGEPITWHEGVSVRNLLHVEDVATATAAL
HNHEALVGDWTPSADRARPGRIETVAASVARQTNPAVPVSVPPENAEANDFR
SDDFDSTFEFTLTGMHVRVPLAEGIDRTVAALISTKE"
12181..22821
/genes="megAI"
12181..22821
/genes="megAI"
/notes="polyketide synthase"

Query Match      4.2%; Score 21; DB 1; Length 47981;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 TTCAGCTCGTGGCGCGGTC 144
Db 22447 TTCAGCTCGTGGCGCGGTC 22467

RESULT 27
AX112026 AX112026 47981 bp DNA linear PAT 01-MAY-2001
LOCUS
DEFINITION Sequence 1 from Patent WO0127284.
ACCESSION AX112026
VERSION AX112026.1 GI:13938925
KEYWORDS
SOURCE
ORGANISM Micromonospora megalomicea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1 McDaniel, R. and Volchegursky, Y.
AUTHORS Recombinant megalomicin biosynthetic genes and uses thereof
TITLE Patent: WO 0127284-A 1 19-APR-2001;
JOURNAL Kosan Biosciences, Inc. (US)
FEATURES
source
1. 47981
/organism="Micromonospora megalomicea"
/mol_type="unassigned DNA"
/db_xref="taxon:47868"
complement (<1..144)
/notes="unnamed protein product; megBVI (megT),
TDP-4-keto-6-deoxyglucose-2,3-dehydratase; SEQ ID NO: 2=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37805.1"
/db_xref="GI:13938926"
/db_xref="REMTREMBL:CAC37805"
/translat="MGDRVNGHATPESTQSAIRFLTRHGGPPATDDVHDWLAHRAAE
HRLR"
928..2061
/notes="unnamed protein product; megDVI,
```

```
TDP-4-keto-6-deoxyglucose 3,4-isomerase,
TDP-4-keto-6-deoxyhexose 3,4-isomerase; SEQ ID NO: 3=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37806.1"
/db_xref="GI:13938927"
/db_xref="REMTREMBL:CAC37806"
/translat="MAVGDRRLRGELQMARGLYWFGANGDLYLSMLSGRDDDPWTW
YERLRAAGRPYASRAGTVVGDHRTAAEVLADPGFTGPPDAARMMOVACHCPAASWA
GPFREFYARTEDASVTVDADWLQRCARLVTELGSRPDLVNDFAREVPLALCTAPA
LKGVDPDLRSEWTSATRVCLDQAQSPQALATEQALDDEIDAVIGRDAAVLVGVV
AELAANTVGNALVATPELPELAALADDPEATRVVTSVTSFGVHLERTAAASDRR
VGVGVPTGGEVTVVAAANDRDPVTDPTDFDVRDGGDAEILSRPGSPRTDLALV
ATLATAALRAAAPVLRSLRSGPVIRRRSRPVARGLSRCPVEL"
2072..3382
/notes="unnamed protein product; megDI, rhodosaminyl
transferase (eryCIII homolog), TDP-megosamine
glycosyltransferase; SEQ ID NO: 4= translated amino acid
sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37807.1"
/db_xref="GI:13938928"
/db_xref="REMTREMBL:CAC37807"
/translat="MRVVFSMAVNSHLFGLVPLASFAQAAGHEVRVVASPALTDVVT
GAGLTAVPGVDGDELVEVHAGODIVEYMTLTDWQDQSHTTMSDMLLGMQTTFTPT
FFALMSPSLIDGMVERCSRWRPDIWVEPLTFAPIAARVTGTPHARMLGPDVATR
ARQSFRLLAHQVEHREDPLAENFDWTLRRFGDPLSFBELVLGQWTDVPIPEPL
RIDTVRTGVGRVYFNGSPVPAWLLKREPRRVCLTLGSSREHGIGQVSGIMELD
AIADIDAEFVATFDQQLVGVSVPANVTAGVPMNVLLPTCAATVHGGTSGMLTA
AIHGVPQIILSDADTEVHAQQLDGLAGLSLPVAGMTAEHLRGAIERVLDEFAVRLGA
ERMDGWRTPDSPAQVVICQDLAADRAARCRQPRRTAEPLPR"
3462..4634
/notes="unnamed protein product; megG (megY), mycarosyl
acyltransferase, mycarose O-acyltransferase; SEQ ID NO: 5=
translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37808.1"
/db_xref="GI:13938929"
/db_xref="REMTREMBL:CAC37808"
/translat="MTVSTNLDTTARPALNSLTGMRFVFAFLVFTFHLVSLRIPNSYV
YADGLDAFWOTGTVGRVSFFILSGVLTWGSARSDSVSWFRRRVCKLFFNHLVTA
AAVLFLDTVGQAVSGEALIPNLLIHAWFPALBISFGINPWSWCLACEAFVLCPLF
LFWISGIRPERLWAAVFAIWAQVPAVADLLPSSPLIPGLEYSALQWFLYTFP
ATRSIEFILGILALITGRWINVGLLPAVLLFPVFFVSLFLPGVYALSSMWLLP
LVLIASGATADLQOKRTFMNRVMVGLDVSFALYVMVFLVIVYGADLLGFSQTEDA
PLGLALPMIIPFLAVSLVSLWLLYRFVLPVNRNWARPASARRKATPEQTPSR"
4651..5775
/notes="unnamed protein product; megDII, deoxysugar
transaminase (eryCI, DnrJ homolog),
TDP-3-keto-6-deoxyhexose 3-aminotransaminase; SEQ ID NO:
6= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37809.1"
/db_xref="GI:13938930"
/db_xref="REMTREMBL:CAC37809"
/translat="MTTVVSYLLEYERADILDVOKVPFASGSLIIGQSVENPETE
YARYHGIAHCVGVDNGTNAVKLESVGVGRDDEVTVNTAAPTVLAIDEGARPVF
YVRDEDYLDMDLVLEAAVPTKAIVPVHLVGCQVMTLRELRLRGLKLVEDCAQ
AHGARRDRLAGTMSDAAAFSFPYTKVLGAYDGGAVTNDDETARALRLRYGMEE
VYVYTRFCHNSRLDEVQAEILRKRLFLDAYVAGRAVQAQRYVLAADLQDSGLEL
PVWTDGNEHFVYVYPRRDEILKRLRDGDISLINSYVFWPVTMTGFAHLGVASG
SLPVTRELAGEIFSLPMVPSPLHLDQRVIEAVREVITGL"
5822..6595
/notes="unnamed protein product; megDIII,
daunosaminyl-N,N-dimethyltransferase (eryCVI homolog); SEQ
ID NO: 7= translated amino acid sequence"
/codon_start=1
/transl_table=11
/protein_id="CAC37810.1"
```

/db\_xref="GI:13938931"  
/db\_xref="REMTREMBL:CA37810"  
/translation="MNSHSTTSSTDAVPEVADIVHDFYHGRKGKGYRAEADALVRVA  
RHTFOAAILLVACCTGSHLVELADSPREVGVGLDLSAAMLATAAENDPGRELHQGDM  
RPSLDRREDVVTMFSSTGILVDELDORAVANLAGHAPGGTLVVEPWPFETPRP  
GWGADLVTSGRRIISRMGSHTPAGLPDRTASRMTIHYTVGSPGAEIHFTEVHVMTL  
FAAAAYEQAFQAGLSCSVGHDLFSGLFVGVAAPGR"  
6592. 7197  
/notes="unnamed protein product; megDIV,  
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmnu  
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID  
NO: 8= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37811.1"  
/db\_xref="GI:13938932"  
/db\_xref="REMTREMBL:CA37811"  
/translation="MRVEIGIEGVFTTPTQFADERGVEGTAYQSDVFVAALGRPLF  
PVAQVTSRRGVVRGVVHFTTTPGSMKATVYCYARGAMDFAVDIIRPGSTFGRAEFV  
ELSAEMVGLYLPVGMGHLFVSLDDTTLYLMSAGVYDPDKRAVHPLDPELALPIPA  
DLDLVMSERDRAVPTLREARDQGLIPDYAACRAAAHVVRT"  
7220. 8206  
/notes="unnamed protein product; megDV, TDP-hexose  
4-ketoreductase (eryBIV, dmnuv homolog),  
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37812.1"  
/db\_xref="GI:13938933"  
/db\_xref="REMTREMBL:CA37812"  
/translation="MVVLGASGFLGSAVTHALADLPVRVRLVARREYVVPVSGAVDYE  
THVDLTPGALAEVADARAVFPFAAQIRGTGWRISDDVVAERTNGLVRDLIAV  
LSRSHAPVVPVPGSTQVGRVTAQVIRDSQDHPGVYDROKHTGQILLKEATAG  
AIRATSLRLPVGPGVAGTADRGVVSIMIRALTQPLTMWDHTGVEVELLYTDA  
AFAVTDALADALAGLHGLTGRSWPLGELVQFQVARSVARHTGDEPVPVSVPPA  
HMPDPSLRSEVPDPARTAVTGVATVMTAEAVDRTVAALAPRRAAAPSPFS"  
complement (8228. 9220)  
/notes="unnamed protein product; megBII-1 (megDVII),  
TDP-4-keto-L-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37813.1"  
/db\_xref="GI:13938934"  
/db\_xref="REMTREMBL:CA37813"  
/translation="MGTGTGASARVRVGRSALHTSLWLGTVPNSGRVTDLRLMD  
HALRGVNCIDTDIYQWRLYKHTBELVGRWFAQGGGRREETVATKVGSEMERVN  
DGLSARHVAACNSLRRLGVHDIIYQTHIDRAAPDEWQAAEHLVSGKGVY  
GSSNLGWHIAAAGESAARNLLGMIHQCLYNLAHRPELDVLPAAQAYGVGFVWS  
PLHGGLSGVLEKLAAGTAVKSAQRAQLLIPAVRPLVEAYEDYCRILGADPABVGLA  
WVLSRPGILGAVIGPRTPEQLDSALRAAELTLGEELELEAIFPAPAVDGFVE"  
complement (9226. 10479)  
/notes="unnamed protein product; megBV, mycarosyl  
transferase, mycarose glycosyltransferase; SEQ ID NO: 11=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37814.1"  
/db\_xref="GI:13938935"  
/db\_xref="REMTREMBL:CA37814"  
/translation="MRLVLTSAFHTHFGQILVPLAWLHTAGHDVVRVASQPELTDVVV  
GAGLTSVPLGSDHLFDISPEAAQVHRVTTDLDFARGPPELSEWFLHGEATSRF  
VFTVYVNDSDVDELVEFAMDRPDLWPEPTFEAGAAKACAGAHAPLLWGSDLTGY  
FRSRQDLRGQRADRPDLPGMLTEVAGRGLDYSEDLAVGQSWDQLESFRLT  
GLESVHTKTLPLYNGSVVFWMLRSDGVRCVFTGGYSALGINSQPEFLRLTLAR  
FDGEIVTTRSGDLPASPDNVLVDVFPNMNLLPFCRAVTHGGAGSWATLHGVQP  
ISVAHEPVCVLRQTEAGLGVFLRDEVDATIMQALATVVEDRSHAENAELRQE  
ALAAPTPAEVVPLEALAHQHRDR"  
complement (10483. 11424)  
/notes="unnamed protein product; megBIV, TDP-hexose  
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12= translated amino acid sequence"

/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37815.1"  
/db\_xref="GI:13938936"  
/db\_xref="REMTREMBL:CA37815"  
/translation="MTRHVTLLGVSGFVGSALLREFTTHPLRLRAVARTGSRDOPPGS  
AGIEHURVLLBEPFRAQVVDIDVVVHLVVAAGSTWRSATVPEASRVNAGIMED  
LVAALRPFPGPAPVLLFASITQAANPAAPSAQHKIEARILQRTEDGVVDGVLIR  
LPAIYGHSGPQGTQGVVTAMIRALAGEPITMMHEGVSVRNLLHVEDVATATAAL  
HNHEALVGDVWTPSDEARPLGEIFETVAASVARQTGNPAVPVSVPPPEANAEANDFR  
SDDFSTERTTLTGWHPRYPLAEGIDRTVAALISTKE"  
12181. 22821  
/notes="unnamed protein product; megAI; SEQ ID NO: 13=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA37816.1"  
/db\_xref="GI:13938937"  
/db\_xref="REMTREMBL:CA37816"  
/translation="WVDVPLDGLTRTHPGPLFPFPLCHNPELRARARQLHAYILE  
GISEDVAVGAALARETRAQDQPHRAVVVASSVTETLTAALAAQGRPHPSVVRGVA  
RPTAPVTVLPQGQAWPGMATLLAESPVFAAAMRACERAFDEVTDSLTELDSPF

Query Match 4.2%; Score 21; DB 6; Length 47981;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTGACTCGCTGGCGGGTC 144  
|||||  
Db 22447 TTGACTCGCTGGCGGGTC 22467  
|||||

RESULT 28  
AY310323/c  
LOCUS 138203 bp DNA linear BCT 06-DEC-2003  
DEFINITION Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene  
cluster, complete sequence.  
ACCESSION AY310323  
VERSION AY310323.1 GI:34766435  
KEYWORDS Streptomyces sp. FR-008  
SOURCE Streptomyces sp. FR-008  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 138203)  
AUTHORS Chen, S., Huang, X., Zhou, X., He, J., Jeong, K.J., Lee, S.Y. and  
Deng, Z.  
TITLE Organizational and Mutational Analysis of a Complete  
FR-008/Candididin Gene Cluster Encoding a Structurally Related  
Polyene Complex  
JOURNAL Chem. Biol. 10 (11), 1065-1076 (2003)  
PUBMED 14652074  
REFERENCE 2 (bases 1 to 138203)  
AUTHORS Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K.J., Lee, S.Y. and  
Deng, Z.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAY-2003) Bio-X Life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China  
FEATURES  
source  
1. 138203  
/organism="Streptomyces sp. FR-008"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:206662"  
complement (574. 1950)  
/gene="fscO"  
complement (574. 1950)  
/gene="fscO"  
/note="putative FAD-dependent monooxygenase"  
/codon\_start=1  
/transl\_table=11  
/product="FscO"  
/protein\_id="AAQ82549.1"

```

/db_xref="GI:34766436"
/translation="MPSTCAPTSGTGRGCPHRRKGTSGRRITAVTPRCVKRKH
QHAQKRWVRGSGVAVVGGSIAGCAALASRGAEITVFERADARLDRGVGIGV
HDARYAELRDAGYNADWAPLNRVWSDRGDAEGRATGTFQFFFRAYNWSLWS
ELRRVPETATYAGAEVAEQADGVTVRLADGEPEFDLVIGQGYRVSREAY
YGAARVAGYIGWGTSGDPSGLPDSGLDAHNTFFGHCMAIRIPDGGSGHRLNWVL
PTGPRDGLHPLDRLTSLPPGRLNAELTEHLRALVAEHPFPFWAAKLLSTPAETTP
IQPIYDLDPHYATDMVLIGDAASVAPPHIGAGSVKALODATALEAAWVAGESKEI
LEGTHAARGVGTAVMGRLARMSAQVEETPDNSAMQAEFDAMWQEQNQSDDRSFG
GGHSIK3R"
complement (2264..3037)
/gene="pabc"
complement (2264..3037)
/gene="pabc"
/notes="putative 4-amino-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="Pabc"
/protein_id="Pabc"
/db_xref="GI:34766437"
/translation="MIEDGEPAGPEALASLALTYHFTTLLVNGEVRGGLDLHLER
LIRDCRTLDALDPDRVKRLRAAPTDGRATVAVTVDFPALNIGNIAADAREGILV
TSRAPDPGPELVRVSRHREDLPEVKSGLCPTLRLRQAQAGYVDLFTGPDGD
ILEGTMWGLVIRGDEVVMPGGEVLGTTQLLRGHRSHRTGPRGRLSRGGLRD
QRRVGVPTGIDDRPEFAHHSVTRLAEIYQALPGSPL"
complement (3150..3818)
/gene="fscRI"
complement (3150..3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="FscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAPADPAGSPERRAAVLADGPDAGAYVACLDPESLTIQVWQ
EFERFGSSSELCSGRCDLIVPSVQOPLMHQFARMDLGKRHRATEVIAVDORITA
STLPNALVAGGTPDVAAILVWMAAEEEGADWAPRKLLSEIDARLDLEGIAA
GVSTPLASRLYLSRQGEYHVTLGRLKLVENRAALVSRAYSVMGLVKGTPWPKVVE
DFIK"
complement (4377..7205)
/gene="fscRII"
complement (4377..7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLLERTTIALVDALGAATGRSSLVLTGPLGTRGSAIIQLRL
SLADRDVVIRLHAAPMEQDFAGVVRQLFETLLGDSPEDARDMDHASFAHR
VLADDAEPGADQAIATAEAVLHGLSLANVASDRLILVDDQLQWSDVPSLRWLY
LAKRLHGLAVVCAIRLDGPRSHHTLAREIREAGTQTLRPSLSLTAHELVRHFHG
EAGDDEFVQACHEASVGNPLFLSLIVGTFGRPLAEASTARRLRPSQLRERLAS
ILRTQAPVYDLAAAILIGEQSDAPTLARLGLDSIGVAGALRALGALRAAPDEP
RTIHSVDRDAESTLTWQORENHEEALLYAAGSPAQVQAQLMAVTVRQWAVD
VLRADATLRGADPTAAGYLCRALDSPAAGVGRLLVELGTAEGRDFPLACERH
LAQWMTLLPEPDRDAVALRIIPTALGPAPLAVDLLRQAEDLGPADGLTGTDRDLA
LLEARLRLCHEDPHELAASARHLKELGPEPPVDTGGERELTIALGLAATLGCPLA
SVVRLAGVLDRPATSARISTPLAIITMAADSVEAVDSWLAVERANERGTPP
ANALVHICQALVHLCGRLAQABOSETALGMEPHQGLDTAAWTHVALESRDP
SLAEKVLRRASIRPQGLTILGLILLEATREIAGDQGTALDSLHSCRGQLEAGWH
NVLFPFRWPAVGVHRRLDGIDITARSALAEVEYARQWGAQVPGVGAALRLGLRDEGG
RGTLGLREAVTVLRGSANQLEALSTLAERLGGGAETALAREATLATSQAPGL
ASRAAGHRSPGPATPDELVTPTERRVVGACRLTQETAGALGVSSRAVEKHLT
HAYRLIGISGRRELIVLFSRAEYAGTD"
complement (7210..10254)
/gene="fscRIII"
complement (7210..10254)
/notes="putative transcriptional regulator"
/codon_start=1

```

```

/transl_table=11
/product="FscRIII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPPIELFLERDRELAALSAVIGELSGRPAVVVTGEP
EQGGLNLRWAAAYATDAGRLVSAHATPAEHYRYGVVQALLAEENRALPRFLTLD
BQGLGLNGLNLAAYADRPTLLVVRDQMLDPSALRWLELBARLLPRAPVALLTSTT
GTAITRPMWSVGTPLTGLATTIELALPELTSGGTATVLRAPGAGDPFAPTEALAEAT
RGIVAVDHDVLDREARAGYSPRADRIDPLRALTAEVVGCDHATRALSDLRDPAADVLR
ALAVCGDLIDPLVCTLAGHSVSESRLRAALASGLTLLRDRHPRVQDAVVRARVLE
EMPAADRAELYARAAGLAQRAVADDQGIADLILLARVPDPAWVDILRRGTSALRGG
RRDLAVAYLARAPLEAEDRARIETGLASVEMVPTAAERERLGLIRATIRPGPGA
GLRATADLCILGGDTTAAARHALAGADSAPAEPRRGPTCAPAVPEAPEASTPTT
HLVPHRPVAGTPPGATGDRRSAAPLDPTGTPRAHDAPRSPGPAAGPDTAGPDEDP
AVERHELTVLCRVASFRLHDDADLDVLPALPDVHNPAVAGVWMEHGVAGVHRE
EVRRLARSAPVPAATGRPLLVVPRLLGCRALLTDDGDEANHGLGALDESHERRAAL
SVAHILTVARLHRRGPDMAADLAAQAEPLDLRLHLPLFTWLALSMITDLQNG
HTERARTEARPTPLLAHESATTAQLLPARGVLARTDDDEPHOAREHFRACRWLLRHG
CANPAVQFWRSLAAEAHTLGDTEEARLVHEVRLARRWGAAAPLGRQAQLSLAVTTE
ENRVENLRAAVTTLSAASATAYTRAVTELAAAEEREENERTTAVAGPAATVLPVAP
PPGPHVPSAVARPPGPRPAVLTGRRTAKAPGHDAVWHGLTAVERTTAALAAQGLGNR
BIATELATTVARLRLSGVYRKLIRGORELRLALVQEAEGS"
complement (10298..13315)
/gene="fscRIV"
complement (10298..13315)
/gene="fscRIV"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscRIV"
/protein_id="AAQ82554.1"
/db_xref="GI:34766441"
/translation="MWLRWAGDRSGAAGPSGAGPSLAPSDARESPVHRRAVP
PAPPLVGRHAQDLVLARHADARAGRPRLVLDGPGIKTKVALLAALAEDEGLAGMT
TLYGSCRAVDATGYSVGRALFGGLVLTVDVHWCDESLRDLWELFLRRAGDGLPLLV
DPGFTSYQLGLYLANLMAADGGLVLTAEVLTGSDTAAVTVYVLAELSTRSLRLENG
LAHRTGGLTAPDALADIVAHHLPASGLGLDPTAEVLAELTAHSPQDPLPSFVRRL
LSHTGSGPLEIVLRLDRSLAGLDGDTGTARLAQAGKVAAVSRVLEHQPFDWVRQ
VARAVAVIGDSDRTVLAALGRVSVVVEEAVLIRLDAGLIHPGHLESHDLVRAAVLD
AVAGSVAVLRRAARLLSDVGRSPEEATQLLLVPGTDDVAVSLRDAQAQAORG
ACEAAALRYERREABKDDVLRLGKALAEITPARSVTLLEHSHLTDIVRAAAT
AVQLGLTCLAVQSGPDGARLFEALDADLAEPEPEPADRELTITTESALLIVGSDDE
KATLPIILRTEGLTPQDTPAQOQLAMLSVLAAEGGAEETVGRARRALRAGV
PLGVMSLLPTSLSLADENEAEBEVLTVLRGSDTAAVTVYVLAELSTRSLRLENG
AVPDMAAQTLAEILGOERWGDATMPHTAASVLTREGDPGRALAEALDKRPHLD
RFVWEYHWTLMARGALAADGLDGLQVFGSCGASMAEAGLTNPVLAIPWLETAALL
GEAGGEEAGRAAAGHTRLAERWGTTRALGYAALARGAARGTARTGALREAVALLAD
SPARGHARGILLGLVEDGAVREGRHLEAVGLRACGCVALARARDELTAAG
GRMREVTASPLDMLTGTERTVAGLVASGAGNEVAESLFTVTVRTVELHLTSYVRKLG
ARRGDLTALREAGATARPAAERRPPHAKRRNP"
13522..14898
/gene="fscMI"
13522..14898
/gene="fscMI"
/notes="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/transl_table=11
/product="FscMI"
/protein_id="AAQ82555.1"
/db_xref="GI:34766442"
/translation="MDSAAPRTLVSLPESGLNPLLVLGALARRRGVLDLWFATDEH
RREEVEALSDVSKVSFASIGEVVPELSATVMDDEVVREVTQSSRKAHRAVVRYP
ALQAKYIRELEQVDEVRPALMVDVCGAVGVDLALARGIPVNLVFPVAVSNVLISHN
PGASYYTPKSPVPNSGIPALMSVQKLANGLTKWRTLGMFLHDMALLRDADAIRK
ELIGAPNPAWTRPVDEAAAVVCSVAELDYPFDIPDRVSLGAVLPLPEAPDDDEVTR
WLDQASSVVMYMGFTITRLTREVAALVEARMSTGHQFLWKLKPEQOHLPEAGSL
PDLRVESVPSQLDVLAHPNVSVFSGHGGNAHEGVYFGKPVQVVRPWVPCFQOAV
RGDRPGISLTLDKPHVDDVDVLDKLTITVTSDDPAFTTEARLGAQRAGRAAADL
VTGLLPAF"
14953..16011
/gene="fscMII"
14953..16011

```

```

/genes="fscM11"
/note="putative aminotransferase responsible for
micosamine biosynthesis"
/codon_start=1
/transl_table=11
/product="fscM11"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTHPVSPALDRELEVSDVSGMISGCPYVRRFEAFAP
EWNGVAGVACSGTALIALNALNIGCPDEVIVFETNVASAMAVTYTGAPFVVD
CGDLDNIDVRIEKEITPRVAVPHVYRGVCDMDALQOINLRVVEVDSAEAHG
VREVGDIACLSFANKIITAGEGVCVLTDPRLAEQAHRAWAFTRFSLHKLKAY
NYETMQAGVADLEIDLEIATREIARYDAGLQDLPGLTLPARDVLWMDLR
ARREELRALHDARGLETFLFFKPMRQPGYLDVFWPTLNARFSDGLYLPHTHTGIL
AAQDEYITGAVRDFYRAG"

```

```

Query Match 4.2%; Score 21; DB 1; Length 138203;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GGCTTCGACTCGTGGCGCG 141
Db 40157 GGCTTCGACTCGTGGCGCG 40137

```

```

RESULT 29
AK110400/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) cDNA clone:002-165-F08, full
insert sequence.
ACCESSION
AK110400
VERSION
AK110400.1 GI:32995609
KEYWORDS
FLI CDNA; oligo capping
SOURCE
Oriza sativa (japonica cultivar-group)
Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

```

```

REFERENCE
1
The Rice Full-Length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-Length cDNA Project Team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T. Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Otono, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN:
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, S., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764

```

```

TITLE
japonica rice
JOURNAL
MEDLINE
PUBMED
REFERENCE
2 (bases 1 to 3161)
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,

```

Namiki, T., Narikawa, R., Niikura, J., Nishikawa, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otono, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES  
source

1..3161  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="002-165-F08"

ORIGIN

Query Match 4.0%; Score 20; DB 8; Length 3161;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCTCGCCGACACC 378

Db 2126 TGCTCGGCTCGCCGACACC 2107

RESULT 30  
AK112081

LOCUS  
DEFINITION  
Oriza sativa (japonica cultivar-group) cDNA clone:001-124-H09, full insert sequence.

ACCESSION  
AK112081

VERSION  
AK112081.1 GI:37988744

KEYWORDS  
FLI CDNA; oligo capping.

SOURCE  
Oriza sativa (japonica cultivar-group)

ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

# REFERENCE AUTHORS

The Rice Full-length cDNA Consortium, National Institute of  
Agrobiological Sciences Rice Full-length cDNA Project Team;  
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,  
Ohtsuki, K., Shiehiki, T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group, Ootomo, Y., Murakami, K.,  
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,  
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,  
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,  
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,  
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;  
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,  
Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,  
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,  
Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice

Science 301 (5631), 376-379 (2003)

22752273

12869764

# REFERENCE AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,  
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayashizaki, F., Hayatsu, N., Hiramoto, K., Hiraoka, T.,  
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,  
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,  
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,  
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,  
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,  
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,  
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y.,  
Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K.,  
Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M.,  
Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Satoh, H., Sakai, C.,  
Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,  
Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y.,  
Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M.,  
Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Tanaka, T., Tomaru, A., Tota, T., Tsunoda, Y., Ueda, M., Waki, K.,  
Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,  
Yokomizo, S., and Yoshimura, A.

Rice full-length cDNA

Unpublished

3 (bases 1 to 3835)

Kikuchi, S.

# REFERENCE AUTHORS

Submitted (12-SEP-2002) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan [E-mail: skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007]

This clone is one of the 32K full-length cDNA clones from japonica  
rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shiehiki, T.,  
Yamamoto, M., and Nakahama, Y.  
FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M.,  
Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M.,  
Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M.,  
Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M.,  
Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,

Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kouda, M.,  
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
Ota, Y., Saitoh, H., Sakai, C., Sakai, K.,  
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,  
Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M.,  
Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Tota, T., Waki, K.,  
Yasunishi, A., and Hayashizaki, Y.

# FEATURES

## source

1. 3835  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-124-H09"

## ORIGIN

Query Match 4.0%; Score 20; DB 8; Length 3835;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCTCGCGACACC 378

Db 722 TGCTCGGCTCGCGACACC 741

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

```

NSHDYWRTHGAAGVGGFRITGASAVLSGRVAYHFGVEGPAVTVDTACSSSLVALHM
AVQALQCEGAMACAGVMGVTVEFVFSRQKGLAPDRCCKAFADGADGTGWSGV
GILLVRLSEARRGHQVLAVRGSAVNSDASNGLTAPNGPAQQRVIRKALAAGLS
PSDVAVEAHGTTTGDPIEAELATATYQDQREOPLNLSGVKSNLGHTRGAAGVAGV
IKVMWRRGVLPRTTLDVDRPSSHWSAGAVELLTEARDWTSQGHPRRAGVNSFGIG
GTNAHVLEVPAPVATPEPPAGFLVPVLLSARSAGLRQGAARGLAGVABVRVPD
VAHALATTAQLDRAVTVLASDRDLRADLAAPFGVVTGTGTPVDGLKAVLFTQGSQW
AGNRELAGTFPFFVRNFAEACHADHILGRAPVREVVFDRLDQTMVTCQALPAV
ETALFVFNASGVYRPLLACHSIELAAHVSGVLDLAEAGELVAARGRLMOALPAGG
MYVACAAEDVAPLDGTVCVAANGPDSVLSGDEADVLAVALAAGRGRTKRLP
VSRFSPPLMEPMGLDPRVAERLTTRAGLTVPVSTLTGKLAALDTPGYVWRQVRAV
RFGDVAELNAGATTYLEGGGALAAALGTILGPEQSCVATLNRKNGAEVPGVUTA
LAEHLVGRVDTWALTDRGTAVGTLPYAFQHERYVGTDETPTSPSPVAARPAE
PVDHVPALVRESVAVLGHDTAAPDLDRFQDLGSLGALIKLNRDLFTGVDPLPS
TLIFDYPNAVLADHLRAGRPAVAAVPRVDSDPEITAVGMSRTRLPFGADSPBE
LWNLVARGDAISGFPVDRGWDLDGLVHPDPAHGTSTYRTGGFLHDAADFAGLFCI
SPREALMNDQOQRLLETSEWALERAGVDPISARGSDVGVFTGTIVHHDYVTRREVEP
DVQGYTMTGTASSVASGRVAYVFGFEGPAVTVDTACSSSLVAIHLAQLRQGECSNA
LAGAGTWMASPDAPLEFSRQGLSADGRCKAYAGADGTGWAEGVGVVVLERLSVARE
RGH"

ORIGIN
Query Match 4.0%; Score 20; DB 1; Length 3900;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGTGATCTCGACTACCC 212
|||||
Db 3022 ACGTGATCTCGACTACCC 3041

RESULT 32
AK121779
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone: J033093K02, full
DEFINITION Insert sequence.
ACCESSION AK121779
VERSION AK121779.1 GI:37991402
KEYWORDS FLU cDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ootomo, Y., Tsunoda, Y.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Ootomo, Y., Ooka, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE 2
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
```

```

Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, Y.,
Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K.,
Nomura, K., Numasaki, R., Ohneda, E., Ooka, H., Ohtsuki, K., Oka, M.,
Ooka, H., Ootomo, Y., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C.,
Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y.,
Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M.,
Tanaka, T., Tomaru, A., Ueda, M., Ueda, Y., Ueda, M., Waki, K.,
Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,
Yokomizo, S., and Yoshimura, A.
Collection, mapping, and annotation of 28K full-length cDNA clones
from japonica rice
Unpublished
REFERENCE 3 (bases 1 to 4116)
AUTHORS Kikuchi, S.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp)
Tel: 81-29-838-7007, Fax: 81-29-838-7007
COMMENT This clone is one of the 32K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cdna/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T.,
Yamamoto, M., and Nakahama, Y.
FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M.,
Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M.,
Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Ueda, M.,
Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M.,
Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohtsuki, M., Ooka, Y.,
Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M.,
Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A., and Hayashizaki, Y.
Location/Qualifiers
FEATURES
source 1..4116
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033093K02"

ORIGIN
Query Match 4.0%; Score 20; DB 8; Length 4116;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 TGCTCGGCCTCGCGACACC 378
|||||
Db 1011 TGCTCGGCCTCGCGACACC 1030
```

```
RESULT 33
AF305325 7080 bp DNA linear BCT 18-JUL-2001
LOCUS Comamonas testosteroni strain BR6020 protocatechuate meta
DICTIONARY dioxigenase locus, complete sequence.
ACCESSION AF305325
VERSION AF305325.1 GI:14861212
KEYWORDS Comamonas testosteroni
SOURCE Comamonas testosteroni
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Comamonadaceae; Comamonas.
1 (bases 1 to 7080)
Proident1,M.A., Mampel,J., MacSween,S., Cook,A.M. and Wyndham,R.C.
Comamonas testosteroni BR6020 possesses a single genetic locus for
extradiol cleavage of protocatechuate
Microbiology 147 (Pt 8), 2157-2167 (2001)
MEDLINE 21387526
PUBMED 11495993
REFERENCE 2 (bases 1 to 7080)
AUTHORS Proident1,M.A., Mampel,J., MacSween,S., Cook,A.M. and Wyndham,R.C.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2000) Ottawa-Carleton Institute of Biology,
Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S
2K9, Canada
FEATURES
source Location/Qualifiers
1..7080
/organism="Comamonas testosteroni"
/mol_type="genomic DNA"
/strain="BR6020"
/db_xref="taxon:285"
141..1538
/note="putative transporter; member of aromatic acid:H+
symporter subclass of major facilitator superfamily"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdK"
/protein_id="AAK73575.1"
/db_xref="GI:14861219"
/translation="MTSQTSTQNTDPTVVNVTVLNEPFGFGFIFLFLIIVLDD
GFDTAIGTIAISGLVHDQIGTIFETITGLIGAMPNAYITLMEYCPDSKRSFTNC
AAVVMGGAISAGLVHDQIGTIFETITGLIGAMPNAYITLMEYCPDSKRSFTNC
MFCGPICGAGGTAANWIPHFWRSLIIVGGVPLILAVLMVLPESVPMVLKN
YSARVRKVRVIAAGTNRARFVLDNGGEAAQAQQPGQGLGVAFVSLKILVTLAW
VTFMGLVIVYLVNMVPLFKAGVPASQAATAALFQLGGVGAIFVGLLMDRWAN
LIIATGYFLTSLGVAFIQVLGGVGLVAVSFIAGLLNAAQSSMQALAEYPTFC
RASGVSMGLGTRFGCIAGSFLVLAELSRRLLELPLVFMVVAIPGVIAALALLKRYA
GGAPRTSMGLGGAH"
1579..2607
/note="putative 4-oxalomesaconate hydratase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73569.1"
/db_xref="GI:14861213"
/translation="MIIDVGHYTTAPALGAWRLQIAGLKDPSTPSVADLKISDD
EICTETNQLRMKRGSDLTIFSPASFMAHHIGDFTSTWAAI CNELCFRVSEL
PDHFI PAAML PQSPVDPATCIPBLVKVQGVNGVNLNPDGSGHWTSPPLSDKS
WYPIYKRVVEYDIPAMHIVSTSCNCHTGTSHYINADTAPMQLTSDLKDFPTLX
FLIPHGGAAPYHWRGFRGLAQEMKPLLEHLNLIYFDTCVYHQPGINLLTEVPT
KNILFASMIGAVRGIDPQTGHYDTRKYITEATQNLTADEKHAYEGNARRVFRLLD
KALKAKGL"
2611..2634
/rpt_type=inverted
2641..2669
/evidence=not experimental
complement(2687..2709)
/rpt_type=inverted
2741..3424
/note="putative 4-oxalocitramalate aldolase"
repeat_region
stem_loop
repeat_region
CDS
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdF"
/protein_id="AAK73570.1"
/db_xref="GI:14861214"
/translation="MYELGVYRNIRQADRAAADGLAALGSAVTEAMGRVGLLKPVM
RPIYAGQVSGTAVTVLLQCDNMHVAEQIQDIIVAATVAECSDGVFGDLLAT
SFQARGARALIIDAGVRDVKLTQEMDFPVMSKAISSKGTIKATIGSVNIPICVAGMLV
TPGDVIVADDDGVVCPFAARAVEVLAQAQKRESFEGEKRAKLASGLVGLDMYKREPL
EKAGLYID"
3603..4520
/note="putative 2-pyrone-4,6-dicarboxylic acid hydrolase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdD"
/protein_id="AAK73571.1"
/db_xref="GI:14861215"
/translation="MSQFETKTPGLDWYANPSKQFKLPAGAVDAHCHVFGNBEFPA
APERKTPCDASAKQLYALRDHLGFARNVVOATCGADNADVADACKVSGKGARVGA
TVKRSISDAELQQLHDAGVRGVRNFARLVADTPKDLMEIAGRIAKLGHVVIYFE
AVDLPDLWDFLTPTTVVDHMGKPDVTGVDSEEFALFLKMRHQNVMSKVSCEP
RLSVTGPKALNGEQAYRDVVPFARRVVEEFDRVLTGTDWPHNPKLDMFDDGLLVD
PIPIAPTAEALQQLLVNDNPMRLYPEV"
4602..5051
/note="alpha subunit of protocatechuate 4,5-dioxygenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdH"
/protein_id="AAK73572.1"
/db_xref="GI:14861216"
/translation="MALEQSKYLDVPGTITFDAGQSKYKYLWLNQFMSLMAKRNRRFR
ANEAYLDWANTEEQKQAVLRLNWCMTGNTIYFLAKI GATDKGFGQAGSMGTG
MTEBEYRAMMNGGGRSAGNRYVGEDGDAQAHQPGSAGNQNRGN"
5055..5924
/note="beta subunit of protocatechuate 4,5-dioxygenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdB"
/protein_id="AAK73573.1"
/db_xref="GI:14861217"
/translation="WARTASVFTSHVPAIGANMDMGKTOEYAWAPLEKGYDFEQWM
KONPDVFLVNDHATAFSLDCIPTFAIGTAEEQPADEGMPRPVKVGVHPDLAS
HIAQSVIQDQDLTIVNMDVHGLTVELSLMGQDPTKTSWPCPIPAFVNVVQYP
VPTGRCFNLRALKAVESYDQDINHVITGTMGSHQLOGARAGLINKEWDNQFLDL
LVENPHGLAQMPHIDYVREAGSEGIELVWMLIARGAMSDVDGPAPLKVAFRHYVFA
SNTAVGHLLLENQ"
6014..6973
/note="putative 2-hydroxy, 4-carboxymuconic semialdehyde
dehydrogenase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PmdC"
/protein_id="AAK73574.1"
/db_xref="GI:14861218"
/translation="MSKTIKVALAGAGAFGINKHLDGKNIIDGVEVSVLVGRFRDQTK
VADKGI AHVATDLAESIALPEVDVILCTPTMAEQAIACMKAGKHVQVEIPLADA
LKDAVEALOKOTGLVAMVGHTRFPNPSHOWHKKI EAGENIQOMDVQTVFFERTN
MNAIQARSWTDLHLWHAHTVDLPAYQAGSPIVKANVQSPITKDI GLIAMDMSIQI
KAANGACTLSLSNFNDGLPTGTFRTYIGDTYILARYDDLYLDEKIDIVSDVDVSMN
GIEIQDREFFFAIREGEPNFSVQGVFCYKYLHDLQQLNAD"
7010..7041
/evidence=not experimental
stem_loop
ORIGIN
Query Match 4.0%; Score 20; DB 1; Length 7080;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 330 GTTCAAGGAGCGCGCGTC 349  
|||||  
Db 1028 GTTCAAGGAGCGCGCGTC 1047  
|||||

MAU66560 7853 bp DNA linear BCT 08-NOV-1996  
Mycobacterium avium EmBR (embR), EmBA (emba) and EmBB (embb) genes,  
complete cds.

U66560  
U66560.1 GI:1619916

Mycobacterium avium  
Mycobacterium avium  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
avium complex (MAC).

1 (bases 1 to 7853)  
Belanger, A.E., Besra, G.S., Ford, M.E., Mikusova, K., Belisle, J.T.,  
Brennan, P.J., and Inamine, J.M.

The embB genes of Mycobacterium avium encode an arabinosyl  
transferase involved in cell wall arabinan biosynthesis that is the  
target for the antimycobacterial drug ethambutol

Proc. Natl. Acad. Sci. U.S.A. 93 (21), 11919-11924 (1996)

97030297  
8976238

2 (bases 1 to 7853)  
Belanger, A.E. and Inamine, J.M.

Direct Submission  
Submitted (09-AUG-1996) Microbiology, Colorado State University,  
Center Street, Fort Collins, CO 80523-1677, USA

Location/Qualifiers  
1..7853  
/organism="Mycobacterium avium"  
/mol\_type="genomic DNA"  
/strain="2151"  
/db\_xref="taxon:1764"  
complement(1..1155)  
/gene="embR"  
complement(1..1155)  
/gene="embR"  
/function="regulator"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="embR"  
/protein\_id="AAC44546.1"  
/db\_xref="GI:1619917"

VSVDALIDGVGAAPVPAARTSIQSYVSTLRILRLSPNPNGLIASVPPGYQLNVAD  
ACDLGRFRGHTAGVOAASRGREFEASGHLAALREWRGPDLDLHNFVDFADANL  
LLERVAHTARAEATACGADGVIQELALVAHPYREPLNAQLITAYVTERQSD  
ALGYRRKTLAEGLDIGPTVKALQQRILRQEPGPGPGQGAARAMLSTHKQST  
VRSEAAVVDNVPVVALRDKAGHYQLNGVTTRIGLDONEIIVLDDTEVSRHHAVII  
DTGDFLITDMKSTNGVQRGRVQRSATLVLDGDIIRIGNSEFVFIRPA"

1156..1332  
/gene="embAB"  
1156..1332  
/gene="embAB"  
/note="divergent"  
1333..4659  
/gene="emba"  
1333..4659  
/gene="emba"  
/function="arabinosyl transferase"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="Emba"  
/protein\_id="AAC44547.1"  
/db\_xref="GI:1619918"  
/translation="MPHDGKQSRQIPRSVAAGIAGLLICLAVPLLPVRQTATVL

gene  
promoter

gene  
CDS

WPQGTVDGHVSQITAPLVSGAPRALDISIPCAVATLPADGGLVWSTLPGGMADGKN  
GLFRANKDWWVAFRDVVAQAORPAVAGACSVLHAWADAGAAGAEFYIGPAAGT  
LPAEKPOYGGITDLKVPAGLSARVIDITRFITAPTVLKOIVLGLTAVLATIIV  
ALAVLDRSRGGTLLNWSPIAWLSYRPGTHLANWRVGLATWADAAVLATLLW  
HVVGATSDGDDYNLTARVAPKAGLYDYRFGTTDAPFDWILGILLSASVSTAGV  
WMLPATLIGICWLLIISHWLRRLGRGGLAANRVAVTAAGVFAAWLPFNNGLR  
PEPLIAGLVITWMLVERALQRLAPAAVVAALLTATLPOGLIAVAALLTGARA  
VQAIRRRASDGLLAPLAAALSLIIVVFRSQTVATVLESARIKYKVGTTIAY  
QWMLRYFTVESNPDSMARPAVLMLCLFGLMLVILLRRGHVPGVAGPWRLLIG  
TVAVGLLLTFTKWAQVQFAGAGLAGALGALTAFACSRIGLHNRNLTLYVTLLE  
VLWATSGINGMYVGNVYDIOPTVITASHVTSFELTSLITGLLAQWHPMDY  
AGTEVDSRRNRVLAETLLVVATIMVGEVASLTGKGVFRFYITTTGKANLAAS  
GLSPKCAMADDVIAEPDANAGMLPQLPGQTFGDPGLGCVNPGKPDGDDLOSD  
PVVTKPGLVNSDASPNKPNVAVSDSAGTAGGKGVNGSHAALPFLDLPARTFVMS  
YGENSLAATATSAWYOLPPRTDRLVPLVSAAGAIWSYKEDGTFYQSLKLOWGTP  
PGSTVPLAEVOPIDIGPOPAENLRPELAWAPEANVARIVAYDNLSEONFAFVP  
PRVTVETVQQLIGSTPVMMDIATAANPCQRPSEHLGVAELPAYRIILPDKQTA  
SSNLWSSSEAGGPFLLQALLRTSTIYLRGWDYRDWGSVEQYFRILVPADAPDAI  
EQGVMTVHGSRQGPTRALP"

4656..7853  
/gene="embB"  
4656..7853  
/gene="embB"  
/function="arabinosyl transferase"  
/note="ethambutol target"  
/codon\_start=1  
/transl\_table=11  
/product="Emba"  
/protein\_id="AAC44548.1"  
/db\_xref="GI:1619919"

/translation="MSVSTVGGDVVRVWVATAGLIGFVLVSVATPLLPVQTTATLN  
WPQGLNSVTAPLISLITFVLDLTATVPCSLVRDLPPGGVILSTGPKKDAALNALF  
VVAHGKRVSDTDRNVVIASARQVAGACSRIEIATSTRAGTATFVGLDPAKPLG  
GGPDPNLRPOIVGVFTDLTGAPAPAGLKSATIDTRFTPTTKLAASMTAILATIV  
ALVALWELDOLGHERMLIPANRFTLADVAIFGVFLVHVLKGAASDDGYILGWA  
RVADAGYMSNYRWFSGSPDPFGWYNLLALMTHVSDASLWMRLPDLFAGIVCWLL  
SRVLTPLPAPVAAASRANWAGVLLTAWPFDNGRPLRPIILRHRVVTGFWPLVAP  
RYSLRTPAALAVITAAFTLGVTGLIAVAGGRPIILRILRHRVVTGFWPLVAP  
MLAAGTILTVFADOTLATVLEATRIATIGPSQAWYENILRYLITPTVDSLSR  
RFGELITVLCFTAVFTMLRRKIPGVARGPAMLMGVIETGTFMFLPTKWHFG  
LPAVAGNAMAALTIVLSPAVLGSNRMAFLAALLPMALCEATUNGWVVSQYVP  
FNSTPKIGGITVYFTFSMFVAALYAIWHLFASRHEGRRLARALTAVPPLAAGF  
MALVFTASWVAGIVRQYTSNADNIRREFSGCGLADDVIVPEPDSNVGPIGGDI  
GLPGGLGGQHPVGSFNPVEHTVAEAIRTPNQPDYDWDAPTKLSAFGINGSVTP  
LPLGLDAARVLAGSYTGAQOSRLTSAMYRLPAPDDGPDVVVTAAGIAGNSVLH  
HHTDGTVLEVYGRPGGDIVPAGRLVVDLYGQPKAWNLRFASDMPADTVAVR  
VVAEDISLTPEDMTAVTPPVEMRSIQEVYSTQPVLMQWAVGLAPCCQPMHVNG  
VTREPRITPDITAKQDITDWDGNGSLGLITDILLRAHVMTSLYLSHDWGRDWGS  
LRRFETIADHAPQALDGLGTATRTGWSPPGPIRKP"

ORIGIN

Query Match 4.0%; Score 20; DB 1; Length 7853;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 TGGCGCTGCCCGGACGCTG 198  
|||||  
Db 2339 TGGCGCTGCCCGGACGCTG 2358  
|||||

RESULT 35  
AE005116  
LOCUS  
DEFINITION Halobacterium sp. NRC-1 section 147 of 170 of the complete genome.  
ACCESSION AE005116 AE004437  
VERSION AE005116.1 GI:10581743  
KEYWORDS  
SOURCE  
ORGANISM Halobacterium sp. NRC-1  
Halobacterium sp. NRC-1  
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
Halobacteriaceae; Halobacterium.  
1 (bases 1 to 12097)  
Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M.,



Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Dale, H., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Spudich, J.L., Isenbarger, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ehardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

TITLE  
Genome sequence of Halobacterium species NRC-1

Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)

205044483

11016950

2 (bases 1 to 12097)

## AUTHORS

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Dale, H., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Spudich, J.L., Isenbarger, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ehardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

TITLE  
Direct Submission

Submitted (14-JUN-2000) Institute for Systems Biology, 4225

Roosevelt Way NE, Seattle, WA 98105, USA

Location/Qualifiers

## FEATURES

source

1..12097

/organism="Halobacterium sp. NRC-1"

/mol\_type="genomic DNA"

/strain="NRC-1"

/db\_xref="taxon:64091"

76..699

/gene="cysg"

/note="synonym: VNG2332G"

76..699

/gene="cysg"

/note="Cysg"

/codon\_start=1

/transl\_table=11

/product="uroporphyrin-III C-methyltransferase"

/protein\_id="AAG20438.1"

/db\_xref="GI:10581744"

GPATADACREAGTVDVVEEYTSAGLVAALGDIDGAEVARSDHGSAVLTGDLAE  
AGAVHEVTVLALTEPAGAGSTETAAAGDLDAALEFSLTVEHWAARERGISSAA  
REGLDATVGAIGPTATETADHGISVDVPPDDAGFEALARATIEDA"

738..2165

/gene="VNG2333C"

738..2165

/gene="VNG2333C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2333c"

/protein\_id="AAG20439.1"

/db\_xref="GI:10581745"

ASTALERAAPHEVVFQSLDDEADTANAGHTVLTDFGSGQDLATHERAGEF  
TFVIADHQPAENAAATYHNLPLEGVGDSLTGAGATYVLARALLGTDATDLAALA  
VVGAVGDRQTVDGELVGNVAADVEAGVLTDTGLALYGTQTRPLRLLEYASEL  
YIPGTDHQGAIRFLDLDVLDHGDGWRTWASLDHDBRRTVWSALVKRAVTVGPS  
AEVTVTLGTYTTLDEEPCALRDASEFATLILNATARYERADVGLAVCMGDRGDALER  
ARSLREERQTLSEGSWTVREGTVTDEDLQWFAHDDRITRETVIGVAGMALSGDETE  
NDRPVFAAKODEPAVKVSARATRVVDAGVDLAVVVEAAAAAGGGGGGHEAAGA  
TYPAGNERAFLDAAGSVLAAQRGG"

complement (2174..6079)

/gene="VNG2334C"

complement (2174..6079)

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2334c"

/protein\_id="AAG20440.1"  
/db\_xref="GI:10581746"  
/translation="WTLTDCAGSPVHVLYVNDDDPAELTRRLKLSAPASGFTVTTVG  
TAALARLDESADCVCYSLSPTDTCIGDILLERLHDHDPPTLFTGRSERTASE  
ATRAGVSYDLPHAGQKSFELGARVRRLTEAAKETVAEMSDRFRILERSIDAVY  
AVDEMRVYNNETMADRDRDAVGVTLWEFFPSLVGTPLEEQYKATMETGVSVS  
SEQLGAPFYVDVRAFPDDDDGLTVFSQETDEREQALALERSETILRHHDVVVVV  
SDAGVVFANAAQRTQSQAPLVGHHLPTIIEDRGSTADATSVSDAISDTLAEL  
SDGTAGFYDNLPIDFDAGTDTTRTLDVRFPTFAADROVLVVRGDTTEQSDVNRHL  
EREDALERIQVVAETELSVDRELEALLAVGCOALGLEIGIVSHIHGEQYTKAHT  
SGVDIAPGEDLATTICESVGDSDVCSFSDATADGRETHPAYBELGLQSYVGAPLI  
VGDRTYGTINFSPATRVADFAEETLELLAEVLVAELSGRDRALQRQKFLFER  
IQETANGVFYFAPSTGSLWSGDVRIHGPYEPFDDAIDAHDPDRETTIAAAV  
DRALDEGEYDLDLRIVRPDGVWRVRAWGERVDSQPQHGARTLRGVIQDITGRKRRER  
EHKALAEYBALLETSGDAIFLLDVTDDDEPRFEARLSPGYEAQTGLETAAKGKT  
PREVVGQSGSEVAANYRCVDEGAPISYRELDVGPDPARFWDTCGLAVVVDGETVRI  
VGIARVTDQVERSLSEATNORLESLEATPLAWMEDPDGTVTRWDEAESMGWT  
REEVVEFPVVPDSQDEFEDHQRVLGGRIKGMELRRRTKSGGDDVLLSLAAVPG  
GADGETTSLAVLEDITPKQLEAKRLAQETREAFDDIPTLQSGDSLAWDAFESAELHYDDLE  
YDLTAVWEHDQSSASLVPLAATGATREAFDDIPTLQSGDSLAWDAFESAELHYDDLE  
HQSDLVASGTPLESGLFVPLGGFVIGVATAGOTFSDTDVDFRILGLSTVAALARA  
TRETLEQNERLQDFASVAHDLRNPVLSVATGFLDIADTGELEYLRDIESAHORMA  
RLIDLLTLASGESITVDACVLDLHAVTTESNGVYDTABATLCYRDDLTYSAGSRL  
TQLFENVRNVAEHGGAATVTVGQGDREGFYVADGDGIFEDERGERGVLDHGVTTSSD  
AGTGFGLSIVEDIKAGHMAVVVTAGADGGARFARFETDP"

gene

CDS

/gene="VNG2335H"

6285..7070

/gene="VNG2335H"

/note="hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2335h"

/protein\_id="AAG20441.1"

/db\_xref="GI:10581747"

/translation="WLYCSGGFGHGDYGNRARGVPRGSPQASTGAPGVRLRS  
QGRSGCTRAQRQDASIGLPPFWGLAPSRARWRLRPTASHKDIPTIQNGPTGRT  
RAVVRTDDGWRTRPSTPKAGSHPPAGVLCAGSGDHLQAGVQTTSMDSLTTPPEEQ  
KYDYFPLETAYKRAFDMMGTVDKLVHAIQQOVLSESEPFYEADSGFSVELPADP  
LDRLAGVQVDADHARDVIDEHTDRICVHLRQFGIGGQADEKA"

7096..7446

/gene="VNG2337C"

7096..7446

/gene="VNG2337C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2337c"

/protein\_id="AAG20442.1"

/db\_xref="GI:10581748"

/translation="MSAESQDGSLEERVDFNFRNFQIOMHGGTAAVQNNINREDGV  
VDLQLGAGSCGGISPMTIQALKSEMTKEIPEIDVVNAGTNGDSSGSGTSPFGEDGD  
TEESADEGPQAPF"

complement (7474..11586)

/gene="polA2"

/note="synonym: VNG2338G"

complement (7474..11586)

/gene="polA2"

/note="PolA2"

/codon\_start=1

/transl\_table=11

/product="DNA polymerase type II large chain"

/protein\_id="AAG20443.1"

/db\_xref="GI:10581749"

/translation="MRPDETYTFETLGLDADFVAEARERGSDPTNVEIPYAKD  
MADRVENLIGDVAERVDRDLDGENSREAALELVADFDVGRVGYDDTDAGLEGAVR  
TAVALLTEGVAAAPTEGIDRVENDNDGDTQYVAVYAGPAGTQAQSLVADY  
ARAMLGIDAFPRDDEIRYAEVDLYSETGLOYSKDAETFTTEHPCVMDGEAT  
GNEEYDFRDLERIDTNSPRGMCVLVAEGIALKAPKIORYTRNDEVAWPMLQDILID  
GIIGADDADETPDAGSSDATDEGDASASTDAEPPRAASDKFLRDLIAGRPVPG  
HESKNGGFLRYGRARNHGNATAGVHPATMHLDDFLATGTQIKTERPGKAGIIPVD  
SIEGTVKLANGDVHRINDPADLVRNGVABILDVGYLVNGYFVENNHLEAPSY  
APEWNIQDLDAAGADVQALRDSPPYVDLTAPTADQAMATDYDAPLHPAYTYLWHDID

VEQFALADADAHTDDDDRGVLVDHTTIVRETLLEALLVTHHQGDDTIRVDDW  
LFLARSLGVSILDEHWEHETLSAAAEWENAVNEVAPFVSQERAPTEIGNMRGPE  
KSEHSLSPAVHLPFPIGAGSGQRDAADAAIAFDMSTPGIEIPVRGDRVCPSCDE  
HYTESRCPDGGTDPHYECRCGAVATPDESRCVPCNGKRDNDLVITQVINDIEY  
HGALRAVGERENAFQOLKAGVLLSAEKTPEPMAGVLRKADHDVAFKDGTVRYDMTD  
LPTVTPAELOVTAQOFRELGVNODIHGDPLEHDDQVLVELVQDVVLSGDAADMLK  
TADVDLLTOYGLDAFVLDLDDRLVGLVFLGMAFPTSAVVGVRVAGFTSASVGYA  
HYFHAARNCPEPNTNWFDESSEWHDPLETLVEARLPDADDEDGALVQAL  
DGVFVSVTEDEETLQRYEAVSRKAPADHLLAVETKGRRLTTPDHSMERWTGCG  
IERVARELTOGADLPAFTQPGDGETSELRSLSLGDGTHPORRGDGSVRTIDVV  
SVEPVSSVDHYSLVATNTLVANGLFTGQDGEDCVMLMDGLINFSPSYLPDK  
RGRMDVMSRIDIPEADIDDEAHNIDIDREYFREFYATRELADPEDVADITLAE  
SVGTDEYTFGCHTHATNIHLGSLSAKVTGLSMMDKMDADOLELARKLRSVAETDV  
ASRVISYHFLPDLGNLRAFSQETCLDCGKRYRMLSGDCRCGGRVNLTVHSGS  
VNKIMTAMRVATEYDCRYTKORLIMRRLSVESDEDTNKGSGISDFM"  
complement (11586..12005)  
/gene="VNG2339C"  
/gene="VNG2339C"  
complement (11586..12005)  
/gene="VNG2339C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng2339c"  
/protein\_id="Vng2339c"  
/db\_xref="GI:10581750"  
/translation="MARVETGGREFVVRDPCSEWTELETAATEDITAAMFTGLG  
AVQDAELYYDQDDFETAAEFPEPLEVACVGNIAITLDGEPFATHATLSRRSGQAL  
AGHLERATVAGELYRELDADLVREHDAATDLDRL"

gene  
CDS

Query Match 4.0%; Score 20; DB 1; Length 12097;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 GAGCGGGCGGCTGGACAC 356  
|||||  
Db 1299 GAGCGGGCGGCTGGACAC 1318  
|||||

RESULT 36  
AE004862/c  
LOCUS  
DEFINITION  
Pseudomonas aeruginosa PAO1, linear BCT 19-FEB-2003  
Genome.  
AE004862 AE004031  
AE004862.1 GI:9950716  
Pseudomonas aeruginosa PAO1  
Pseudomonas aeruginosa PAO1  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 17589)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
Wong,G.K.-S., Wu,Z. and Paulsen,I.T.  
Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
20437337  
10984043  
2 (bases 1 to 17589)  
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,  
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,  
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,  
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,  
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,  
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,  
Hancock,R.E.W., Lory,S. and Olson,M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,

University of Washington Genome Center, University Of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 17589)  
Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)  
Direct Submission  
Submitted (04-FEB-2003) Department of Molecular Biology and  
Biochemistry, Simon Fraser University, 8888 University Dr.,  
Burnaby, British Columbia V5A 1S6, Canada  
This represents the February 3, 2003 version of the continually  
updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,  
from pseudoCAP (see http://www.pseudomonas.com for latest updates  
and links to alternate annotations). PseudoCAP is coordinated by  
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert  
E.W. Hancock (University of British Columbia, Canada). We welcome  
submission through www.pseudomonas.com of any proposed changes.  
'Protein name confidence' is used to rate our confidence of the  
accuracy of the protein name.  
Class 1: Function experimentally demonstrated in P. aeruginosa.  
Class 2: Function of highly similar gene experimentally  
demonstrated in another organism (and gene context consistent  
in terms of pathways its involved in, if known).  
Class 3: Function proposed based on presence of conserved amino  
acid motif, structural feature or limited sequence similarity  
to an experimentally studied gene.  
Class 4: Homologs of previously reported genes of unknown function,  
or no similarity to any previously reported sequences.  
-----  
Location/Qualifiers  
1. 17589  
/organism="Pseudomonas aeruginosa PAO1"  
/mol\_type="genomic DNA"  
/strain="PAO1"  
/db\_xref="taxon:208964"  
complement (75..569)  
/gene="mreD"  
/locus\_tag="PA4479"  
complement (75..569)  
/gene="mreD"  
/locus\_tag="PA4479"  
/note="Protein name confidence: Class 2 (high similarity  
to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="rod shape-determining protein MreD"  
/protein\_id="AAG07867.1"  
/db\_xref="GI:9950717"  
/translation="MARATQGSNNVWVWISLLVGLVAVAPMPTFTTIGRPLMLML  
LTVWLLPVRGVMTAWLLGLAQDVLVGNLLQNALILGLITFTVLSLHQRMRFFPA  
WQCCVLVWVYGLAQLQLMLNALTGNRPPTLFLPALVSALLWPVYALLQFVRLR  
LNVR"  
complement (569..1561)  
/gene="mreC"  
/locus\_tag="PA4480"  
complement (569..1561)  
/gene="mreC"  
/locus\_tag="PA4480"  
/note="Protein name confidence: Class 2 (high similarity  
to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="rod shape-determining protein MreC"  
/protein\_id="AAG07868.1"  
/db\_xref="GI:9950718"  
/translation="MLAVLSVLMVVDARFDYLEPVRSKLGNLTTPFVGLAEMPVRAM  
EGVDRQSSRELSIAENRKAESLLMORRVOKLAUTEQNVRLKELNSAALVDVDKV  
LVSELIGVDNPFQIRIMDKGENDGVFGQFVLDASGLMGVVEVMEPYTARVLLITD  
TTHSPIVQVNRGLRAVAGTGNPERLERLYVADTADTKEGDGLGQFFAGYP  
VATYKEVHDSQPPAVVRVAVTPAKMRNSRVLLVFSRSTPEQRANDAAQAQEAADK  
KAAQAQAPQAQAPAAAPSPATPAQAQAQAPAPAPATQPPAAPAANGRR"  
complement (1641..2678)

REFERENCE  
AUTHORS  
CONSTRM  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source  
gene  
CDS  
gene  
CDS  
gene



```

GSSMAVVGCAVMAATPGVVFDFSRQALAAADGRSKAFGAGADGFGFSEGVTLVLLER
LSERARNGHEVLAVRGSALNQDSNGDSLSAFSGPAQRVIRIQALSSCGLERFGDVDAV
EAGHTGTALQDTEANALDITGDRDADRPILWLSKSNIGHTQAAQVGLLKVVL
ALRNGELPATLVEETPTVDVSSGALLAGNQWRGRTERRARVSASFSGTNAH
VIVEAPERHEHTTHDQRPVPLVVSATTAALRAQAQIAELLERPADLAGLGLG
LATRARHREAAVASTREBAEVAAGAAATADAVGVTEVDGRNVVFLPFGQ
GSOWAGMELLSSSSVFAGKIPACDESMAQWDMKVSVDLQAPGAPGLDRAVDVQP
VLFANVSLAEIWRSGVGPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSLMRS
LSGEGMAAVALGEAAVRRLRRLRQRLSVAAGNPRSVVSGEPGALRAFSDCHAE
GVRIDVDVYASHSPQIRVREBELLETTDIAAPRVTFHSTVESRMDQFELDAR
GIRVDDVYASHSPQIRVREBELLETTDIAAPRVTFHSTVESRMDQFELDAR
LHRDGDLSAFRLSMATAVHSVDIEMVHPVVQVVEEAEEADGAEDAVVGS
YWRNLRETVRFADAVTRIAESGYDAFIEVSPHPVVQVVEEAEEADGAEDAVVGS
LHRDGDLSAFRLSMATAVHSVDIEMVHPVVQVVEEAEEADGAEDAVVGS
ASDELAVYSVQPIEKPESGNLDGMLVVTLELSPWTEMLCEAINANGRALRCVD
TASLREMAQAQVAACTGPRGVLSLLSSDESACRPGVPAGVGLLTLVQALDAGVDA
PWMLTQGVARTPADDLRAQTTAHGFAQVAGLELPGRWGVVDLPESVDDAALRL
LVAVRGGRAEDHLAVRDRHRRVVRASLPQSGRSWTGHVTLVLTGAASPVGDQ
LVWLDRGAERLVLGACFGDLLAAVEEAGASVVCQAQDAALREALGDEPFTALV
HAGTLTNFGSISEVAPPEAETIAAKTALVLDEVLGDAVEREVYCSVAGIWGGA
GMAVYAGSAYLDALAEHRARCSCTSVAMTWMALPGGAVDDGYLRERGLRSLADR
AMTWERVLAAQSPVAVADVDPVLSSEGPAATPTALPAELAGRGQAEEPDSCPT
GEPALAGLSLDEQENLLELVANVAEVIGHESAEINVRRAFSELGLSDLSNMAAL
RKLSTAGLSLDEQENLLELVANVAEVIGHESAEINVRRAFSELGLSDLSNMAAL
FLTDAADFPGFGPIITPREALAMPQORLMELETAWEAERAGIDPDALRGITDGVFGG
MNGOSYMOALLAGAEABVDGQVGLGNSASVLSGRIAYTFWEGEPALTVDACSSSLGI
HLAMQALRGECSLALAGGVVMSDPVTFVDESTQGLASDGRCKAFSAPADGFALSE
GVAALVLEPLUSRANGHVIVIRGSANQDASGLAAPNGPSQERVIRIQALASG
VPAADVDEHAHGTGELGDIIEAGALITATYQDRDRPLRLGSVKTNIGHTQAAQAA
GVTKVLMHAGMLPSLHABELSPHIDWESGAVEVIREVPPWAGERPRAGVSFG
VSGTNAHVIVEEAPAEAAARTGRPLFPVLSGRSEAAVAAQARALAEHLRDTPELGI
TDAATLATGRAPDVPEAAVILGDRAGVCAELDALAEGRPSADAVAPVTSAPRKPVLV
FPQCGAQWGMADLLESSVFAESMRCAEALSPHTDWKLLQVDRDGGDPPHERVD
VLPQVLFISVMSVLSGASRDELETVLARQGRVAAVAVNGPSTVWAGTFAELDEFFAE
VIRELDDQGMVSGVMSRDELETVLARQGRVAAVAVNGPSTVWAGTFAELDEFFAE
AEAREMKPRIVARYASHSPSEVARIEDRLAAELGTTITAVRGSVPLSTVTVGTVDISA
MDASVYWRNLRLPFAQVRLVBOGFTFVEVSPHPVLVMAVEETAHAGAEVTCV
PTLRREGSGPHEFLNLLRAHVGVGADLRPAVAGGRPAELPTVPEHQRFWPRPHRP
ADVSALGVRCGHEPILLAAVDPVGHGAVTGLSTDEQPLAEHVVGRTLVPGSVL
VDLAAAGEVDGLVPLELVLQRLVLVLAGALLRMSVGPDESGRKTIIDVHAEDVA
DLADAQMSHATLAQVAAAGPRDTEQWPPEDAVRILPDDHDYDGLAEQGYEYGPSFQ
ALRAAKMDSDSVAEVSIAADEGYAFVPLLDVAQTLISLIGALGEPGGKLPFAWNT
VTLHAGATSVRVVATPAGADAMALRVTDPAHLVATVDSILVSTGEKWEQEPGRGG
EGELHALDMWGLABGSGTRVVAADASDLDAVLSGREGPEPDVAVLYVEPGDDPRAAA
RHGVILAAALVRLWLEBELPGATLVLTATSGAVTSDSDSVPEPGRGAAMGVLRCAOA
ESDPFVILDTADBPGLMPVAVPDPQLALRGDDVVFRLSPFLAPSALTIFAGTQRLVP
GDGAIDSVAFEPAPDPOPLRAGEVRVDRVATGVNFRDLVLLALGMLPKQADMGTEAAG
VVTAVGPDVDAFAGDRLVGLFQGAFAPIAVTDRHLARVDPDGMSDADAAPVTAYTT
AHVALHDLGLRAGOSVLIHAAGGVGMAVALARAGAEVLA TAGPKHGTALRALGL
DDHEIASSRETGFARKERTGRGVVDVNLNLTGELLDESADLLAEDGVFVEMGKTD
LRDAGPRGRYAPDLGEAGDDRLGEILREVWGLLGAELDRLPVSAWELGSAPAAQ
HMSGRHVGLVTLQAPAPVDPDGTGLITGGTGLRLLARHLVTEHGVRLHLLVSRAG
ADAPGSDLRABIEDLGEASABIAACDADTADALSDLLDGLPRPLTVGVHAGVLAADGL
VTSIDEPAVEQVLRKADVAWNHLELTANTGLSFVLFSSAASVLAGPQGVYAAANE
SLNALALRRTRGLPKALGMLWAQASEMTSGLDRIARTGVAALPTERALLAFDSA
LRRGEEVFPPLSINRSALRAAEFVPEVLRGMVRAKLAAGQAQEAAGNVPVLAGRSE
SDQVAGLAELVRSHAAVSGSADQQLPERKAFDLPDLSLAALVELNELLGTAVGRL
PSTVLFPHPTLPAVHRHRLRFAFPAVDLGRDLDELEKALREALSLEDGDDHVGQRL
ESLURWNSRADAPNSTSAISEDASDDDELFSMLDQRFGGGEDI"
19. .4470
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of module 3"
97. .1482
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoacyl ACP synthase
of module 3"
1693. .2670
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyltransferase domain of
module 3"
3406. .3921

```

```

/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoreductase domain
of module 3, possibly non-functional"
4171. .4428
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyl carrier domain of
module 3"
4471. .10722
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of module 4"
4471. .5847
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoacyl ACP synthase
domain of module 4"
6054. .7026
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyltransferase domain of
module 4"
7165. .9216
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of dehydratase and
enoylreductase domains of module 4"
9433. .9984
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of beta-ketoreductase domain
of module 4"
10225. .10483
/misc_feature
/genes="eryA"
/notes="putative"
/function="approximate span of acyl carrier domain of
module 4"
10723. .20235
/misc_feature
/genes="eryA"
/notes="ORF3 encoding modules 5 & 6 for
6-deoxyerythronolide B formation; putative"
/codon_start=1
/transl_table=11
/protein_id="AA026495.1"
/db_xref="GI:152695"
/translators="MSGDNTWEEKURRYLKRTVTLELSDVTLRREVEHRAPIAV
GMACRFPDVSPEFVFYSGGDAIAEAPADRWGEPDPAELGMLAAGDPDAGF
FGISPREALAMPQORIMLEISWEALERAGHPVLSRGSATGVTGTVGYGRPDE
APDEVLYGTGTASSVAGRVAYCLGLEGPAMTDTACSSGLTALHAMESLRDEC
GLALAGVTVSSSGAFTFERSOGLEADGRCKPFSKAADGFLAEGAVLVLORLSA
ARREGRPVLAVLRSVNOQASNGLTAPSPACQORVIRALENAGVAGDGVDEAH
GTGRLGDPLEVHALSTYGAERDPPDPLWIGSVSGNIGHTQAAAGVAKVIALR
HEMPRTLHDEPSPQIEWDLGAVSVVQARSAPAGPRRAGSVSSFGISGTNAHVIV
EEAPEDEPPAPSPGPVPLVLSGRDEQAMRAQAGRLADHLAREPRLSDTGTLLAT
RRSAWHRVAVVGRDRLDALAGLAVADGRIADRTATQARTRGVAVVFPQCGAQWG
MADLPLESOFVADSIFDCERALPHVDWLSITDLSGARPLDRVVPVOPALFAVWVL
AALNRSGVPAVAVGHSQGEIAAAHVAGALTLEDAKLVAVRSEVLRLGGQGMAS
FGLGTEQAERIGHRGFAGALSIVAGNSPRSVVAGSGPGLDELIAEACGTAARRIPV
DYASHPQVLESLEELTELAGTSPVSADVALYSTTQOPIDTATMDTAYWRLNRQ
ADPHATRLQALAEAGDAFVSVPHVLTGIEATLSDALPADACAGVGTTRDRDGL
RFQDARTQLAEAGDAFVSVPHVLTGIEATLSDALPADACAGVGTTRDRDGL
RYQVWTEAEWESASLAGVLLVLTGCPVPSELSDAIRSGLEQCATVLTCDVSRSTI
GTALAEADTALSTVSLRDEAVDPDLALVQALGAAGVEAPLWLTNRNAVQV
ADGELVDPAMVGGGLRVGIEQPGRWGVVDLPESVDDAALRL
AIRADGKIVARLVPAPARAARTSRPVTLVGTGCGIGAHVRLWLRSAEHLVLL
GRGADAPAGSELRELTALGTGTIAACDVADRLEAVLAEAREGRTVSAVMHAA
GVSTPLDDITLAEATEIADVKVGTVNLDELCPDLDAFVLFSNAGVWGSGLASY
AANAFLDGFARRRSESGAPVTSIANGIWAQNMAGDEGGYLSRQGLRMDPRAVE
ELHITLDGQTSVSVVMDRRRRFVELITAAHRLPEIETAGARAEARQSESGPALAQ
VLTLSAEAREHRLAHLIRAEVAVLGHGDDAIDDRAFDLGFPDFTADPLRNLAA
VTVREAAATVVDHPTITRLADHYLERLVGAAEQAPALVRFKADDDPIALVGMA
CRFPGGVHNPGLWEFTVGRGDAVTEMTDGRGWDLDALFDPPDQRHGTYSRGHAFLD

```

GAADFDAAFFGISPREALAMPQORQVLETTWELFENAGIDPHSLRSGSDTVFLGAAAC  
QYGVGDVAVPEDESEGVLLTNGSSAVSVGSVAVVGLGEPVAVTVDTACSSILVALHSAC  
GSLRQDGLAVAGVGSVMAGPEVTFEFSRQGLAVDGRCKAFSAEADGFGFAEAGVAV  
VILRLSDARRAGQVLGVVAGSAINQDASNGLAAPSVAQORVIRKAWARAGITGA  
DVAVEAGTGTRLDGPDVBEASALLATYIGKRSRSGSEVLLGSVKSNIHQAAGVAGV  
IKVILGNRLGVPMLCRGRSEPLTIEWSSGVLELAEVSPFPADGVRRAAGVSAFGV  
SGTNAHVITAEPEPEPLPGFVGLAAANSVPVLLSARTETALAAQARLLESAYDD  
SVPLTALASALTGRAHLPRRAALLAGDHEQLRGQRAVAGVGAAPGATTGASAGGV  
VFVPGQAGMEAGRLLSVPVFAESIAECDAVLSEVAGFSASEVLEORPDAPSLEL  
VDVQVPLFSVMVSLRWLWAGAGVSPSIVIGHSQGEIAAAVAVAGVLSLEDGVRVVALR  
AKALRAGAGKGVSLAAGPGERARALIAWEDRISVAANSPSSVSVVSGDPRALAEILY  
ARCDEGVRKATLPVDYAGSRHREVEIRETILADLDGIGARRAAIPLYLTGHERDGG  
ADMPFRYTHNLRSQVRFEPAESAADVADGTFVEMSPHPVLTAAQAEIALDAVATGS  
LHRDTAEHLIAELAEARAHVGVADWRNVFPAAFPVALPNYPFEPORYWLAEVSDOL  
ADSRVYDWRPLATTTFDLEGGFLVHGSAPESITSAVERAGGRVVPVVASADREASAAAL

Query Match 4.0%; Score 20; DB 1; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTCGACTCGCTGCGCGCGGT 143  
Db 10354 TTCGACTCGCTGCGCGCGGT 10373

RESULT 38  
LOCUS AR049368 20235 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5824513.  
ACCESSION AR049368  
VERSION AR049368.1 GI:6005407  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20235)  
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.  
TITLE Recombinant DNA method for producing erythromycin analogs  
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;  
FEATURES  
Location/Qualifiers  
source  
1. .20235  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 4.0%; Score 20; DB 6; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTCGACTCGCTGCGCGCGGT 143  
Db 10354 TTCGACTCGCTGCGCGCGGT 10373

RESULT 39  
LOCUS AR095529 20235 bp DNA linear PAT 08-SEP-2000  
DEFINITION Sequence 3 from patent US 6004787.  
ACCESSION AR095529  
VERSION AR095529.1 GI:10023473  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20235)  
AUTHORS Katz, L., Donadio, S. and McAlpine, J.B.  
TITLE Method of directing biosynthesis of specific polyketides  
JOURNAL Patent: US 6004787-A 3 21-DEC-1999;  
FEATURES  
Location/Qualifiers  
source  
1. .20235  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 4.0%; Score 20; DB 6; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTCGACTCGCTGCGCGCGGT 143  
Db 10354 TTCGACTCGCTGCGCGCGGT 10373

RESULT 40  
LOCUS SEERYABS 20444 bp DNA linear BCT 28-FEB-1992  
DEFINITION S. erythraea eryA gene for 6-deoxyerythronolide B synthase II & III.  
ACCESSION X62569  
VERSION X62569.1 GI:46977  
KEYWORDS 6-deoxyerythronolide B synthase; acyl carrier protein; acyltransferase; beta-ketoacyl synthase; dehydratase; enoyl reductase; ketoreductase; multifunctional subunits; thioesterase.  
SOURCE Saccharopolyspora erythraea  
ORGANISM Saccharopolyspora erythraea  
Pacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.

REFERENCE 1  
AUTHORS Bevitt, D.J., Cortes, J., Haydock, S.F. and Leadlay, P.F.  
TITLE 6-deoxyerythronolide B from S. erythraea: cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme  
JOURNAL Eur. J. Biochem.  
REFERENCE 2 (bases 1 to 20444)  
AUTHORS Bevitt, D.J.  
TITLE Direct Submission  
JOURNAL Submitted (19-SEP-1991) D.J. Bevitt, Biochemistry Dept., Univ of Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK  
COMMENT For related sequences see X56107 & M63677.  
FEATURES  
Location/Qualifiers  
source  
1. .20444  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL 2338"  
/db\_xref="taxon:1836"  
/clone="SE55, DE10, pBK25"  
/clone\_lib="genomic in lambda EMBL3"  
37. .42  
46. .20268  
/gene="eryA"  
46. .10749  
/gene="eryA"  
/notes="ORF B"  
/codon\_start=1  
/transl\_table=11  
/product="6-deoxyerythronolide B synthase II"  
/protein\_id="CAA4448.1"  
/db\_xref="GI:581651"  
/db\_xref="GOA:Q03132"  
/db\_xref="SWISS-PROT:Q03132"  
/translation="MTDSEKVAEYLRRTATDLRAARQIRELESDFPIAVSMACRLPFG  
GVNTPQRLWELLREGGETLSGPTDRGMDLARLHHPDPNPGTSYVDKGGFLDDAAGE  
DAAEFGVSPREAAAMPQORLLETSTWELVENAGIDPHSLRSGSDTVFLGAAAGV  
DTAAFGVSESVTVGAPAVASGKISYTMGLEGPSISVDTCSSSLVALHVAESLRR  
GESSMARNHEVLA VRGSAINQDASNGLAAPSVAQORVIRKAWARAGITGA  
LSEARNGTALGDPTEANALDITYGRDADRPLWLSGKSNIGHTQAAAGVTLGLKVV  
ALRNGELPATLHREETPHVDSSGGVALLAGNQPWRCERTERRAAVSAPGISGTNAH  
VIVVEAPERHRETHADGRPVLPVVSARSTALRAQAQIAELERPDADLAGVGLG  
LATPARHERHAAVASTREAVRGLRETAAGAATDADAVVEGVTEVDNRNVVFLPQ  
GSQWAGMGAELSSSPVFAGKACDESMAPMDQWVSDVLRQAPGADGLVDVVPQ  
VLFVAVMSLAELWRSYGVFAAVVSGSQAIAAAHVAGALTLEDAAKLIVGRSLRMS  
LSGEGGMAVALCEAAVRELRPWDRLSVAANGPRSVSVSGEPALRAFSDCAAE  
GIVRQIDVDVYASHSPQIERVRELETTGDIAPRPARVTFHSTVSRSDGTFLDAR  
YWRNLRRETVRFADAVTRIAESGYDAFIEVSPHFVVVGVVEEAVEEDAGEDAVVGS  
LHRDGLSAFLSMATAVHSGVDIRWDVLPGAAPFALPTYPFQRRYVLPQAPAA  
ASDELAYRSWTPIEKPESGNLDGMDLVVVTPLISPEWTEMLCEAINANGRALRCEYD

TSASRTMAQVAQAGTGRGVLSLLSDSACRPGVPAGVGLLTILVQALGDGAVDA  
 PWCLTQGAVRTADDLAPACTTAHGAQVAGLPGRWGVVDLPESVDDAALRL  
 LVAVLGGGAEADHLAVRDLGRVVRASLPQSGSRSWTGHVTLVTCGASPVGDQ  
 LVWADRAERLVLGACPGDOLLAVEREAGASAVVCAQDAALRALGDEFTVALY  
 HACTLTFNFSISEVAFEEAETAAKTALLAVLDEVFGDRAVEREYVCSVAGIWWGA  
 GMAAYAGAYLDAALAEHHRARGSTSVANTFWALPGHVDGGLRGLRSLSDR  
 AMTWERVLAAQSVAVADWVPLSEGFAATPTALFAELAGRGQAEADPSGPT  
 GEAPOLAGISPEQOENLLEIVANAVAEVLGHESAAENVRFAFSELGDSLNAMAL  
 RKLSASTGLRLPASLVPDHTPTALAQHLRLBLVGDQAQAVRVGADESEFIALV  
 GICRPFPGGISEQWVRVLAEGANLITGFPADRGMWDIGRLYHPDNDPNTSYVDKGG  
 FLTDAAITLQAGRPDREALMDPQORLMLTAEWAVERAGIDPDALRDTDTGVFG  
 MNGQSYMQLLAGAEVRYQCGHNSVLSGRIATYFWEGLPALTVDTCSSLVGI  
 HLAQALRRGECISLAGGVVMSDPTFYFDQFSGLASDGRCKAFSARADGFALSE  
 GMAVLEPLPSRANRANGHVLAVGSAVNDGASGLAANPFSORVROALASG  
 VPADVLEVEAGHTGELGDIIEAGALIAVGGODRDLPLRGVSKTNGHTQAAAGAA  
 GVTVLVMRHMGLPSRLHADELSPHIDMESGAVEVLRSEVPWAGERPRAGVSSFG  
 VSGTNAHVVEEAPAEQEAARTERGPIPLVLSGRSEAVAAQALAEHRDTPELG  
 TDAWTLATGARPVDRAAVLGGDRAVCAELDALAEGRPSADAVPTAPRKPVLV  
 FPQCAQWGMARDLLLESSEVFAESMSRCALSPHTDWKLLDVVRGDGDPHERVD  
 VLOPFLFSIMVLSAEIWRAGHVTAAVUGHSQESI AAHVAGALSIAAAKVVVALRSQ  
 VLRDLDSQSVLSAGSRDELETVLAEWDCGRVAAVNGPCTSVIAGPTAELEDFPAE  
 AEAREMPKRVISAGSRSEVARIEDRLAEGLTITAVRGSVPLHSTVTVGEVIDISA  
 MDASVYRNRLRPLVFPQAVRGVLEQGFDTFVEVSPHPVLLMAVETAEHAGAEVTCV  
 PTLREQSGHEFLRNLLRAHVGVGADLPVAGVPAELPTYPFQHFWRPFRPP  
 ADVSALGVRAEHPILLAAVDVPGHGAVFTGRLTDEQPLAEHVVGRTLVPGSVL  
 VDILAAAGEDVGLPVEELVLRPLVLGAGALIRMSVGAPEDESVRTIDVHAADVA  
 DLADAQWOSHATGLAOGVAGPDETPOMPEDAVRIPLDHVDGLAEQGVYGPSFQ  
 ALRANWKDSVVAEVSIAADEGYAHPVILLDAVATLSIGALGEGGKLPANWT  
 VTLHAGTSVRVATPAGDAAMALRVTDPAHLVAVTSLVVRSTGEKWEQEPREGG  
 EGEHLADWRLVAFEGSTRVVAADSDLDLAVRSPEPEDAVLRYEPEGGDPRAA  
 RHGVLMAALVRLWELELPGATLVATISGAVTSDDDSVPEPGAAAMGVTRCAQA  
 ESPDRVILDTADPGMLPAVDPNPQALRGDDVFPRLSPALPSALTLPAGTQLVBP  
 GDGADSVAPTEAPQPLRAGBVRVDVATGVNFRDVLILALGMYPKQKMDMTEAAG  
 VVTVAGVDPAFAPQDVLGLFOGAFAPVTDHRLARVDPDGSDDADAAVPTAYTT  
 AYHALDLGRACQSVLIIHAAAGGVMAVALLARRAGAEVLATAPAKHGTALGL  
 DDEHIASSTREFGAKRFGGVVDVNLVSLGELLDSDALAEADGAVFVEMGKTID  
 LRADGDFRYPALPDAGDGLTEILREVGLLGGELDLPLVPSAMELGSAPALQ  
 HMSGRIHVKLVTOPAPDPDGLVLTGGTGLRLRLHVLTVHGVHLLVLSRRG  
 ADAPCSDELRAEIDEGASAEIACDADTADRALSALLDGLPRLTGVHAGVADGL  
 SLNALADRVQGLPAKALGWLMAQSEMTSGDKRIARTGVAAALPTERIALALFDSA  
 LRRGGEVFFPLSINRSALRAEFYEVILRMVRAKIRAAQGAQAEAPNVDRLAGRSE  
 SDQVAGLAEALVSHAAAVSGSADQLPERKAFKDLGDFSLAAVELRNRLGTATGVRL  
 PSTLVFDPHTPLAVAEHLRDLFAASPAVDIGDRLDELEKALEALSAEDGHDHVDGQRL  
 ESLLRWNRSRBDAPSTSAISEDASDDLEFSLMLQRFQGGEDLI"  
 139. .1416  
 /gene="eryA"  
 /product="beta-ketoacyl synthase"  
 1471. .2763  
 /gene="eryA"  
 /product="acyltransferase"  
 4261. .4455  
 /gene="eryA"  
 /product="acylcarrier protein"  
 4513. .5787  
 /gene="eryA"  
 /product="beta-ketoacyl synthase"  
 5839. .7116  
 /gene="eryA"  
 /product="acyltransferase"  
 7117. .7563  
 /gene="eryA"  
 /product="dehydratase"  
 8743. .9234  
 /gene="eryA"  
 /product="enoylreductase"  
 9418. .10164  
 /gene="eryA"  
 /product="ketoreductase"  
 10315. .10509  
 /gene="eryA"  
 /product="acylcarrier protein"  
 10750. .20268

/gene="eryA"  
 /note="orf A"  
 /codon\_start=1  
 /transl\_table=11  
 /product="6-deoxyerythronolide B synthase III"  
 /protein\_id="CAA44449.1"  
 /db\_xref="GI:46979"  
 /db\_xref="GOA:Q03133"  
 /db\_xref="SWISS-PROT:Q03133"  
 /translation="MSGDMGTEELRRYLRKTRVTLELDSVTARLEVEHRRAGEPIAIV  
 GMACRPGVDSPESFWEFVSGGDALAEAPKRWEPDPDARLGLMGAAGPDAGF  
 FGSPREALAMDQQRIMLEISWEALERAGHDVPSLNGSATGVTGTVGVYDGRPE  
 APDEVLYGTGTAGTASVAGSRVAYCIGLEGFAMTVDTCSSGLTALHLAMESLRRDSC  
 GLALAGVTVMSFGAFTFERSQGLAADCKPFAADGFLGAEAGVNLVLRDSEAH  
 ARREGPVLAVLAGSANNQDNGSLTAPSGAQQRVIRRALENAGVAGDVYEAH  
 TRTGLDPLFVHALLSTGABRDPPDPLWIGSVKSNIGHTQAAAGVYWKAVLALR  
 HGMPRTLHPDESQLENDLGAUVSVQASRWPAGERPRRAGVSSFGISGTHAVTV  
 EEAPEADPEPPADSGEPLVLSGRDEQNRQAAGRLADHLAEPNLSLDTGFTLAT  
 RASAMBERAVVGDREALAGRAVADRADRTATGQGENSRVAVVFPQGAQWQ  
 GMARDLIRSEVPADSTIRDCERALAHPVMSLTDLLSGARPLDRVDVQVAFVWVS  
 LAALMRSHQVEPAAVVHSGQETAAAHVAGALTLEDAALVAVRSVRLRLGGQGM  
 SFLGTEOAAERIGREFAGALSIAVNGPRSVVVVAGSGPLDELIAECEAEHAKRI  
 PVDYASHSPVESIRRELLTELAGISVYSADVALYSTTGQPIDTATMYANLAR  
 EQVRFQATQLAAGADAFVEVSPHPLVTVGLTEALDLSALPADAGACVGTLRDSG  
 GLADPHTLGEAYAQGVEDWSPAFADARVELPVPFQRYWLPITPTGGRRADDDDD  
 RYQVWREAEWESASLAGRVLLVTPGVPSELSDAIRSGLEQSGATVLCDESRSTI  
 GTALEADTALSTVGVAAPHGEAVDPDLALVQALGAAGVAEALVWLTNRNAVQ  
 ADEGLVDPQAQMVGLGRVVGIBQPGRWGGLVDLDVDAASIRSLAALADPRGEEV  
 ALRAGDKVAVLPAPARATHPLEPLAGTVLTGTGGTGAHLARLARSAGSHLV  
 LGRGADPAGASRLRELTALHGTGTAAQADARLEALAEALAEAGRTVSAVMH  
 AAGVSTPTDLDLTEAEFTIADVKVGTNVNLDLCPDLDAFVLFSNAGVWGSFGLA  
 SYEALHITLDGFGTASVVDNRFRFELTAARPLFDEIAGARASARQSEEGPDLA  
 QRLAALCLCDKREHIAHLIRAEVAAVAGHDDDAIDDRADFRLGDFDSTAVDLNR  
 LAATVQREAAVTVEHPHTIRLADHYLERLVGAEEAQAPALVREVPKQADPILAI  
 VMACRPPGGVHNPGELEWETVGGGDVATVEMPTDRGMDLALPDQPHGHTSVSRHG  
 AFDGAADFAAFPGISPREALAMDPOORVLETTWETTFENAGIDPHSVRGSDGVFL  
 GAAYQGVQDAVVPEDSEGVLGNSAVSVAYVVLGEGPATVLTACSSLEVAL  
 HSGACDGLDCGLAGAGGVVAGPEVTFEFSRQGLVADGCKAFSAEADGFLPE  
 GVAVQVQLRSLDGPAGRGVGVVAGSAINQDQATNGLAAPQAQVQVIRKAWARA  
 GITGADVVAEAGHTGRLGDPVEASALLATYKSGSGSVLLGSVKSNIGHAQA  
 GVAGVIVKVLGNLGVPPMLCRGERSPLIWSGGGVLELAEVSPWPAPAGVRAAGV  
 SAFVGSQDAVPIIIEPEPEPEPEPGVGVVLAANSVPVLLSRTALTAALAEALLE  
 SAVDSVPLTALASALATGRAHLPRRAALAGHEQLEGRQLRAVGAEGVAPGATGTA  
 SAGVGVVFPQGAQWEGMARGLLSPVFAESIACDVAI SEVAFSAEYLEQRPDA  
 PSLEVRDVQVPLFVSVMLARLWAGACVSPSIVIGHQSEIAAAVAVGVLSDEGVR  
 VVALKALRALAGKGGVSLAAPGERARALIAPEWEDRI SVAAVNSPSSVVSQDPEA  
 LAELVACEDGVRKATLPVDYASHSRHVEIRETILLDLGLSARAAIPLYSTLHG  
 ERDMGPVRYWYDNLRSQVRPDEAVSAQSPDGHATFVMSPHVLTAAVQEIADAAVAL  
 GSLHRDTAEELHIAELARAHVGHVAVDMRNVPFAAPFVALPNYPPEPQVLAPEVSD  
 QLADRSYRVDMWRPLATTVDLEGGFLVHGSAPESLTSAVEKAGGVFPVVASADRALA  
 ALREVPGEVAGVLSVHTGAANALAHOSLAGEAGVRAPLMLVTSRAVGLSEPEVDEQ  
 AMWGLGVWGLETPERWGLVDLPASPGDGEAFVACLDGHEQDOVARLDRHARY  
 RRLVRAPIGTRSSWEAGTALVTGGTGLGHHVHRLARGVEDVLVSRGVADPA  
 AAELAEALGPKTITITACDADRBEQKULEELRGQRPRVTVVHTAGVPSRPLH  
 AAELEALGPKTITITACDADRBEQKULEELRGQRPRVTVVHTAGVPSRPLH  
 VAILGHSSPAVQDQDPFTELGFDSLTVGLRNQOATGLPATLVEFHTVRL  
 ADHIGQQLSDGTPAREASSALRDGVRQAGSVRSYLLDLAGLSDPREHFDGSGFS  
 LDLYMDADGPGEVTVICAGTAIIISGPHHEFTRLAGARGIAPVRAVPPQTEEGEFLP  
 SSMAAAVQAQDAVIRTOGQDFVYVAGSAGALMAYALATELLDRGHPRPGLVDVY  
 PPGQDAMNWLBELTATLFDRETVRMDTTLTALGAYDRLTGTQWRRETPGLTLLVS  
 AGEPMGPWDDSMKPTWPFEDHTVAVPGDHTWQEHADA TARHIDWLGGGNS"  
 10867. .12111  
 /gene="eryA"  
 misc\_feature

Query Match 4.0%; Score 20; DB 1; Length 20444;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 124 TTCGACTCGTGGCGCGGT 143  
 |||||

```

Db      10381  TTCGACTCGTGGCCGGGT 10400

RESULT 41
E38021/c
LOCUS   E38021      31422 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38021
VERSION   E38021.1 GI:18626910
KEYWORDS  JP 2000245457-A/2.
SOURCE    Streptomyces avermitilis
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS   Omura,S. and Ikeda,H.
TITLE     Avermectin aglycon synthase gene
JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT   OS Streptomyces avermitilis
            PN JP 2000245457-A/2
            PD 12-SEP-2000
            PF 24-FEB-1999 JP 1999046961
            PR
            PI SATOSHI OMURA,HARUO IKEDA
            PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH Key Location/Qualifiers
            FT CDS (1)..(14643)
            FT CDS (14824)..(31419).

FEATURES             source
     source           1..31422
     /organism="Streptomyces avermitilis"
     /mol_type="genomic DNA"
     /db_xref="taxon:33903"

ORIGIN
Query Match      4.0%; Score 20; DB 6; Length 31422;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      366  CCTCGCCGACACCGCACCG 385
          |||||
Db      27210  CCTCGCCGACACCGCACCG 27191

RESULT 42
BD097650/c
LOCUS   BD097650      31422 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION A method for procuding avermectin derivatives.
ACCESSION BD097650
VERSION   BD097650.1 GI:22643224
KEYWORDS  WO 0162939-A/2.
SOURCE    Streptomyces avermitilis
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS   Endo,H., Yamaguchi,H., Kanda,Y., Hashimoto,S., Omura,S. and
            Ikeda,H.
TITLE     A method for procuding avermectin derivatives
JOURNAL   Patent: WO 0162939-A 2 30-AUG-2001;
            KYOWA HAKKO KOGYO CO LTD,THE KITASATO INSTITUTE,HIROFUMI ENDO,
            HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI HASHIMOTO,SATOSHI OMURA,
            HARUO IKEDA
COMMENT   OS Streptomyces avermitilis
            PN WO 0162939-A/2
            PD 30-AUG-2001
            PF 23-FEB-2001 WO 2001JP001381
            PR 24-FEB-2000 JP 00P 47405
            PI HIROFUMI ENDO,HIROYUKI YAMAGUCHI,YUTAKA KANDA,SHINICHI PI
            HASHIMOTO,

PI      SATOSHI OMURA,HARUO IKEDA
PC      C12N15/54,C12N15/53,C12N9/10,C12N9/14,C12N1/21,C07C327/30, PC
C12P17/08
CC      A method for procuding avermectin derivatives FH Key
FT      CDS (1)..(14643)
FT      CDS (14824)..(31419).

FEATURES             source
     source           1..31422
     /organism="Streptomyces avermitilis"
     /mol_type="genomic DNA"
     /db_xref="taxon:33903"

ORIGIN
Query Match      4.0%; Score 20; DB 6; Length 31422;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      366  CCTCGCCGACACCGCACCG 385
          |||||
Db      27210  CCTCGCCGACACCGCACCG 27191

RESULT 43
MSGY222/c
LOCUS   MSGY222      41156 bp      DNA      linear      BCT 03-DEC-1996
DEFINITION Mycobacterium tuberculosis sequence from clone Y222.
ACCESSION AD000010
VERSION   AD000010.1 GI:1702968
KEYWORDS  Mycobacterium tuberculosis
ORGANISM  Mycobacterium tuberculosis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE 1 (bases 1 to 41156)
AUTHORS   Du,L.
TITLE     Direct Submission
JOURNAL   Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
            Beaver Street, Waltham, MA, USA, 02154 du@ctic.com
COMMENT   GSDB:S:1004711
FEATURES             source
     source           1..41156
     /organism="Mycobacterium tuberculosis"
     /mol_type="genomic DNA"
     /db_xref="taxon:1773"
     /clone="Y222"

ORIGIN
Query Match      4.0%; Score 20; DB 1; Length 41156;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      92   TCGGCCCGCGCCGAGGGTTC 111
          |||||
Db      14176  TCGGCCCGCGCCGAGGGTTC 14157

RESULT 44
SFU78289
LOCUS   SFU78289      43280 bp      DNA      linear      BCT 13-AUG-1997
DEFINITION Streptomyces fradiae tyactone synthase, starter module and modules
            1-7, (tylg) gene, complete cds.
ACCESSION U78289
VERSION   U78289.1 GI:2317859
KEYWORDS  Streptomyces fradiae
SOURCE    Streptomyces fradiae
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 43280)
AUTHORS   DeHoff,B.S., Sutton,K.L. and Rostek,P.R. Jr.
TITLE     Sequence of Streptomyces fradiae tyactone synthase gene tylg

```



JOURNAL Unpublished (1996)  
REFERENCE 2 (bases 1 to 43280)  
AUTHORS DeHoff,B.S., Sutton,K.L. and Rostock,P.R. Jr.  
TITLE Direct Submission  
JOURNAL Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate  
Center, Indianapolis, IN 46285, USA  
COMMENT tylg.  
FEATURES  
source  
1. 43280 Location/Qualifiers  
/organism="Streptomyces fradiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1906"  
1. 43280  
/gene="tylg"  
816..14234  
/notes="ORF1; putative"  
/codon\_start=1  
/transl\_table=11  
/product="tylactone synthase starter module and modules 1  
& 2"  
/protein\_id="AAB66504.1"  
/db\_xref="GI:2317860"  
/translation="MSSALRRVQSNCGYGLDLSNTAQTQDQEDVDGPDSTHGGE  
IAVVGSGRCPGAAGVEEWEILRSRGMTQDDGTWRAALDHAGFDAGFFGMNAR  
QAAATDPOHRLMLELWEALDAGIVPGDLTGTDTGVAGVADDVAVLTRASVAG  
GYTAGLHRLAANRLSHFLGRLPSLVVDSASISIVAVOLACESLRGETSLVAG  
GNLILTESTYWMERNGALSPPGRCHTFDARANGVYRGEGGAVLKPDLAALADG  
RVYCVTKGGANNDDGASLTPDREAQAVLRQAYRRAGVGTGAVRYVLEHGTGTRA  
GDVEAAALGAVLAGADSGRSTPLAVSGVKTNVHLEAGAGIVGLIKATLVIRKEL  
VPSLNFSTPNPDIPLDRLRVTQTEQWNEEDDRPRVAGVSGFGMGKTNHLVIRKEL  
DVAATWADRSRGGSGAGSIVSGVVPVSVGRSVVREAAAGRLAEVAGVGGLA  
PAAAGSSGAGSGAGSIVSGVVPVSVGRSVVREAAAGRLAEVAGVGGLA  
DVAATWADRSRGGSGAGSIVSGVVPVSVGRSVVREAAAGRLAEVAGVGGLA  
GGAGAGGAGAGAGGGGTVLPQCGTQVCMGAGLIGSSSEVFPAASMRCEARALSV  
HVGWDLLEVVAGSAGLVRVDPVPTVMVMSLARYQWAGVDVAVVAGHSGQSEIRAA  
TVAGALSDEAAAVLRAAGLIGRYLAGRGMAAVPLPAGEVEAGLAKWPGVGAAN  
RSPRVVCSVAGEQPEVPCQAEQVQRLIPVDYASHSHRVHEDLKEGELRVISGRIP  
VLAHQTAEAADRSVHNGTLRROODSPHRLTSTAFAWANGATLTWDPALPPGHL  
TLTPTFPNNHLDITLIDGGGDATQKESGLPRLELRLGLPSSQKQLFDLIDYCR  
HTAVLGLDTAAEVDPLSKKQIQSMGTVELRNLLTGTGLATPILVLDTPTRPA  
LAQFLHTLIDSGSGSVLAPAKQSFQAQPIAVVGMGCRFPGGSGFPEALMRLVWEG  
VDVSPPPGDRGVGLYDPEPGVAGKSVYREGFLHDAEFDAEFPGISPREAVAM  
DQORLILTSWEALBRAGIDPHSLHSGRTGVAGVMPQVGPRLAEGAGSGDGLLT  
GTSQSVSGVAVTLGLEGPVTVDTACSSSLVALHVAQALRGCECDMALAGVTVM  
AGPMVFERRQGLADGRCKAFADGADGTANAGAGVILVERLSARLGHPEVLAV  
VCGAVNQDGSNGLTAPSGSQSVIRQALGNARLTADVDDVVEAHGTGRIGDPIE  
AQALLGTGRDGRFPVWLGSKSNIGHQAAGVAGVTKQVLMARYGMLPRTLHVD  
EPKSHVDWAGGVWLLTEAREWPGVDRPRRAAVSAFVSGTNAHLILEAPDPAESA  
TTPVRSVSSAAVLIDARSQVFPVVSGRSVRVREAGRLAEVVEAGVLDABEAVT  
MAGSRFGVRAVLARCEAEELAGRLALAGGDPDAGVTVGAVDVPETGSGGGVVLVE  
PQGTQVWGAGLIGSSSEVFPAASMRCEARALSHVENVDDLELVVSGGAGLERVDVOP  
VTWAVMSLARYQWAGVDVAVVAGHSGQSEIRAAVAGLSDLEDAVAVLRAGLIGR  
YLARGMAAVPLPAGEVEAGLAKWPGVVAANGPASTVVSDDRRAVAGVQVFEAG  
GVQARLIPVDYASHSHRVHEDLKEGELRVISGRIPVCSVAGEQPEVPCQAE  
YWFNLNRNRFSAVGVGLKEGHRRTFVESAHPVLVHAIPTQAEADRSVHATGTLR  
RODPSPHRLTSTAFAWANGATLTWDPALPPGHLTTLPTFPNNHLDITLTPPTAT  
TTQSTPDARVRYTWKALTSSVPVPHSIGLCLVAPPTDGLLGLTTLVLSERGS  
VARLEVPFGARRAEVALLKESMBSEAGENTTVSLGLGVLESTDAVTSIALLOQVSD  
IGVFAARVWALTTRAVAVPGETPDQAGLWGFGRVALELPDIWGLLIDPAVDEL  
TRTPSQPTPPTPQTPNRRALALEAAVLAGRDEGDQVAVRAGSIYGRVRAEA  
AGAAWSPQSVLITGCMGALGRILARLAAAGERLVTTSRRGPPEAPGAELAEALR  
GHGCEVHAACDVARDALAAVTPVAPNAVFTAGLDDAVIDTLPSPESPTVRGAK  
VCGAELLQLDITKGLDAPVLFSVVTGWNAGCGAVAAANRALDALAEERRAAGLP  
ATSVAGLWGGGMAAGAGEESLSRGRUAMDPPDAVALLGAMGRNDVCTVVDVDW  
ERFAPATNAIRPGLFDTPPEAREALTAAAGTSATPDGAPELARLRLSMINETERLKL  
VELVRETAARFHPNTDAIGAERPFKSGFSLTSLRLNRNAGTGLKLPATVIFD  
HPSPTALARLILDRITGAPAPADEPLPVAVADDPPVTVGMACRFPGGAGPEA  
LKLWTEEDVIGAPTDRGWLDSDVPDPEGVAGKTVYREGFLHDAEFDAEFGLI  
SPREAVAMPQORLILTSWEALBRAGIDPHSLHSGRTGVVGLTHQYASRLHAPE  
EYGGYLLTGKAGSVVSQISYTLGLEGPSSLIDTACSSSLVALHNAQAALRGCECDMA  
LAGVTVMAAPGLFVEFGRIADGRCKAFADGADGTANAGAGVILVERLSARLDMAR  
LGHPVLAVVCGSVANNQDGSNGLTAPSGSQSVIRQALGNARLTADVDDVVEAHGTG

TELGDPIEAQALLGTGRDRAECPPVWLGSKSNIGHQAAGVAGVTKQVLMARYGM  
LERTLHVDEPSRHWDSAGSAGVRLLEAREWPGVDRPRRAAVSAFVSGTNAHLILEAP  
EALALEATDEAPEAPADVDTVTEALEAPADVTEAREGAKAPGSPPEAQFAGVVP  
VVSGRSRVVVREAAAGRLAEVAGVGLADVAVTMAGSRFGYRAVVLARAGEAELAG  
RLALAGGDPDAGVTVGAVDVPETGSGGGVLPVPGQCTQWGMGAGLIGSSSEVFAA  
SMRECARALSHVENVDDLELVVSGGAGLERVDVQVTVWAVMSLARYQWAGVDVAAV  
VGHSGEITAAATVAGALISLEDAAVVVALRAGLIGRYLAGRGMAAVPLPAGEVEAGLIA  
KWPGQVAVANGPASTVVSDDRRAVAGVQVCOAGVQARLIPVDYASHSHRVHEDLKEG  
ELERVLSGIRPSRPRVPCVSTVAGEQPEVPCQAEQVQRLIPVDYASHSHRVHEDLKEG  
HRRFTEVSAHPVLVHAIPTQAEADRSVHATGTLRROODSPHRLTSTAFAWANGATLT  
TWDPALPPGHLTTLPTFPNNHLDITLIDGGGDATQKESGLPRLELRLGLPSSQKQLFDLIDYCR  
DSGFRVAGRLAGSEHDLTDAVSGVTLTPTQAEADRSVHATGTLRROODSPHRLTSTAFAWANGATLT  
PLVLPAGGVRVQMSVSAADSESGRRRVAIHSAPAAVHSAAGGDSAGVWTRHGEGTL  
VPDPEPTPDADNARAWPAGRPFAELIYRFGALYGEAFAGAFAGVAVWQPDALL  
AEVLLPDRASTAGGFGFGLDAAALQWIIAGLLEVEDPAVLLPFQWQVSLXYATG  
AGALRVLLTKAGDGVSLQADTSGAAVLSLGAALVNRPLARRKLDVLLGTLDAGERSLY  
RVWQPRLLPAGPSPRSWAVLGPDADELACTPGIGDQDQGPCTALYEVREARLAKAAG  
APRPEAVPLVSGAGATPESVRQTRTERCLTALQDWLDAEELVDTPLIYLTRGAAVAV  
PQBEITGLACAGVWGLVRSARSHPGRFALVDTGHPDDRTALPLAIRAVLDGAGQLS  
LRSRGRAAGADELAELAGLEAEPFAACDAADREALARLAEVADPPLTGVTHA  
LRAQTARTVLLRAGTPERQGFAPDPACTVLTGATGTIGRLAARLAEVADPPLTGVTHA  
AGVLDGDTDLTTPERTGMKADALNLHELTRTSPLSVFAVSGAAGILGRPCQ  
ANYAANTFEDALAQHRAHGLPAYSLSANGLWGGATGTMTHLSGTDLRRMRSGIAPM  
THDQGLALFDALAAASREDPLVPMKLDLALVREAREHGPDAVPGPLGLLGPAPAV  
RQAAFPVCGGAPAGGEGTAERLAGEERLURELRLVRAEVSGVGLSGDPAVEP  
GRPPKDLGFDLSLTAVELNRNLAGATGLRFTALVDFRTPSQAQVAYLAELAGFRDGG  
DTAAAFEGLEALAAVAGALAEEDLDRDLRRRLRLTALAAALTPQGRNPSAPAPASDL  
DERLDSANDDLFAFTEEQL"  
942..3929  
/misc\_feature  
/gene="tylg"  
/function="encodes starter module"  
942..2156  
/misc\_feature  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase KS-Q of  
starter module"  
2571..3957  
/misc\_feature  
/gene="tylg"  
/function="encodes acyltransferase domain of starter  
module"  
3675..3929  
/misc\_feature  
/gene="tylg"  
/function="encodes acyl carrier protein domain of starter  
module"  
3993..8471  
/misc\_feature  
/gene="tylg"  
/function="encodes module 1"  
3993..5264  
/misc\_feature  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase of module 1"  
5631..6617  
/misc\_feature  
/gene="tylg"  
/function="encodes acyltransferase domain of module 1"  
7410..7949  
/misc\_feature  
/gene="tylg"  
/function="encodes beta-ketoreductase domain of module 1"  
8220..8471  
/misc\_feature  
/gene="tylg"  
/function="encodes acyl carrier protein domain of module  
1"  
8541..13970  
/misc\_feature  
/gene="tylg"  
/function="encodes module 2"  
8541..9812  
/misc\_feature  
/gene="tylg"  
/function="encodes beta-ketoacyl ACP synthase of module 2"  
10260..11246  
/misc\_feature  
/gene="tylg"  
/function="encodes acyltransferase domain of module 2"  
11319..11876  
/misc\_feature  
/gene="tylg"  
/function="encodes dehydratase domain of module 2"  
12861..13415  
/misc\_feature  
/gene="tylg"



```

/misc_feature
/function="encodes beta-ketoreductase domain of module 2"
13719..13970
/gene="tylg"
/function="encodes acyl carrier protein domain of module
2"
CDS
14351..19945
/gene="tylg"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"
/protein_id="AAB66505.1"
/db_xref="GI:2317861"
/translation="MTAEIVATKRLGALREAREPIAVVAMSCRYPGVTTTPEDLMR
LLADERDAVSLPRDGDWLDYDPOGGPSTYAREGGFLSHCAGFDEAFEGISPRE
ALAMDPOORLLETWEALERAGVTADRARSRTGCVAGVWYDDYGARVLYGAGGPP
EDLEGVYVNGSAGTASGRVYTFGLRPAVTVTACSSLSVLHLAVRALRNGECMD
ALAGATVLTPTVDFSRORGLAPDRCFAFADSDGTWAGAGMLLIQRLSDAR
REGREVLAVIRGSVNDGASNGLTAPNGRAQOVIEDALRDAGVDPDVAEAGHT
GTELDPIEAGALLATYGTARTERPLWLSKSNIGHTQAAAGVAGVILKVLAMRHG
RLPRLTHVDREPTVDMKGVRLITEPVPMFGEAGEPRRAGVSGFSGAGTNAHVYLE
SPAGEPPAAGPEPTDGGAMTVSGRGPAAALRAQARLYLDGTGTGGQAGQAGP
GTAEVAGLAHARTAFRRHLYRHPVFARALDEVCALPHEPLRLDMFAPGPEA
TALFSGOSORAGSGRLRYRHPVFARALDEVCALPHEPLRLDMFAPGPEA
EPLDRTEPTOPALFALONALFELAEHGLRAEALCGHSVGBITAAHAGVLTLPDAAR
LYVAGRLMQALPAGGAMAALRAVAEEIAPLPERPAGELALAAVNGPSSVVSGDEA
VLELQWRAGREARRILAVASHAFSPRDMGLTQFDVARTILFPAPTILVSLTIG
TEVTEETICTADHWVQAREPVRLDMARTLRADGIDTFVELGPGVLSMARDCAD
RPDGTTCAGGCTPDPLTLPLLRSPVETGDAEPGCFERALATAYAHGVPLIAP
APDAASAAVAELPTYATQTHWLDPAAPALPAGLDAGHPLLSAALDPPGGRIT
VMTGASALT.PMAADHSVHGRTVLPGLTALLDLALHAAAPRVGELTFEAPLVLPEDGEV
RLRVLPADPAGSVRELSSHGEGGWTRHATVLTGTTTTTPAGPAGPAAAPPGD
APDLAEYERFADAGIGYFAFRGLASRWDRDAILLVRLPSELAGEADRYGLHP
ALLDLAAHGAHGAHGMPLFTWNGVTLHARGAHLKRVLTTPAGDPAVATVADPA
GRPFTASLTLPVTYQTLTAAEARAPLYRVWRTGLPDTGTATDHTWAGGPGDL
LPGETPHIPDLALADGTATPFRVLADLRGYGTATPRELSQLALVQWADAEEA
EGRIVLVTRAVIDGVTDPAAATWGLVRAAQSEHPCCFALLDTPDSRSPRLPR
VAGTAEQLARDGTLTLLASPLTEATIPAGARLPALDGTGLTGTGTSIGAEARHLVTR
HGARKLLTSRGPQAGAEVLVLAELGAHADVAACDVADRALLRALLDRVPAGHP
LTAVLHTAGDGGVLTATQTPORLAALVRKADAVRNLUHEUTQGHLSAFILYSSAAG
VLGSAGGYAAANAYLDSFAVWRSSRLGPAVSLGFWPDGGMASGLGGTDTLARR
Query Match 4.0%; Score 20; DB 1; Length 43280;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; spH gene, spI gene, spII gene,
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 169 GCCACGGGCTCGGCTGCC 188
|||||
Db 13893 GCCACGGGCTCGGCTGCC 13912
POL505006 48000 bp DNA linear BCT 06-JAN-2003
Sorangium cellulosum spig gene (partial), spH gene, spI gene,
spII gene, spik gene, spil gene and spiz gene.
POL505006
AJ505006.1 GI:27528333
cytochrome-P450-monooxygenase; OH-methyltransferase; polyketide
synthase; regulatory protein; spig gene; spH gene; spil gene; spII
gene; spik gene; spil gene; spiz gene.
Sorangium cellulosum
Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1
Knauber,J., Bloecker,H., Mueller,R., Nordsiek,G. and Beyer,S.
The spirangiene synthase from Sorangium cellulosum So ce90 - Module
duplications as strategy behind the evolution of polyketide
synthases
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 48000)
Beyer,S.

```

```

TITLE Direct Submission
JOURNAL Submitted (06-AUG-2002) Beyer S., Nbi-Mx, German Research Centre
for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, GERMANY
FEATURES
Location/Qualifiers
source
1..48000
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/db_xref="taxon:56"
1..7558
/gene="spig"
<i>.7558
/function="biosynthesis of spirangiene"
/codon_start=2
/transl_table=11
/product="polyketide synthase"
/protein_id="CAD3448.1"
/db_xref="GI:27528334"
/db_xref="COA:Q8GBX7"
/db_xref="SPTREMBL:Q8GBX7"
/translation="ILGSDARLASALEAAGAHVTRHRDPGAWIDAIRGGGAPEVVAVL
ACTTRDPAPSGALATGCAADADALAOQAAERELASCALVWTERAVATPDDVDVL
DLVHAPVWGLVRRAQVENPDRDALIVDDADDAASLPALVAAVAAADASEPOLA
LREGRIIVRLRIGRGDDALVPSSAAAKLHVAAKGTLESALVTHEENTRPLAGSQI
RVAVRAGLFRDADALDMYPGDPGLGAGAGVLEVPGVTELPAGDRVNVHAAAG
AFPTIALDHMRVSRIPAGLSFVEAASIPVVLTYAYGLVDRGRPERVILHAAAG
GVGTAAVOLARHLGAEVFATASPPKWTALRALGLDDDDHTASSRGDFFAHVFRTSGGR
GVDTVDSILARELVDAISLVALAPDRKFMVKSDVRDPLAAASHPSVAHYRFPDLA
EAGERIQOMALAEVALFERGVLPVPTPCDLRPAQAFRALAORALGVKVLITPR
ADPRSVCTGDTGRTGALVARHILVGHVRSVLVSSRRGPAAQCADLRELEAAG
ALSAVACDAGDRALERLALIPADRPLTAVVHAAGTDDGVLTDGGLTSLTPRERUWSA
AVSLDGYWAEERTGLTGLGFLVLSGLVSGVSPQYAAANTFIDLAHQRSRGLP
KVDAAHDELTPLDLSAGFVLGASGLVSGVGGVGGVGGVGGVGGVGGVGGVGGVGGV
PDPVALAHAVAPITLGLVRAGARRKLANNAASAPLAORLAGLAPKPERAMIDV
VPELGVVGLPPSHLGDGRPLKELGDSLMAVELRNLAAI VGTPLPATLLEDYPT
VRLPGLATLALVGRGDAAPSAARASAGASDAAIALVAMACHPPGCVLPEDLFR
LIREGGDASIAFFPENRGDLDALSDGARQGGFLHDADLDAEFFGISPRETIALDPQ
ORLLIEMTWEALERAGIIPASLHGSKSGVGVFGVSGVYVRLHGPFGSQAQWESURGY
IGTGSASVAGRIAYALGLEGPALSVDTACSSILVHLACHAIRRSGCSLALAGGV
TVMPTGIFGSGAOGHAGAPGRCCKFSADANGAGWSEGAIVLLRERLSDKNGHPV
LAIVRGSVNDGKQGLTAPNGRAQOVIRDALASAGIAPADVDAVEAHGTGTTLGD
PIEARLAFATYGAHAKERPMLGSLKSLNGHTQAAAGIGGVVKNVILAQSVLPRTL
HAASPSQHIDWSSGAILLTTEVPWPKGSPRRAVSSFGISGTNAHVVLEAPRAAA
GDQAAEPARAPAPPVVLISAKDAALRAQAKHLHAHAEHEDVALVDVAASLATT
RSHFERAAVADDAALISALALAEGRAGAGTAVAGEALPPGKVAFFVFGGQSWPS
MARALLASPAFRATEACERALAPHDVMSLLAVLGGDEHAAPMLRVDVVPVLFA
VMIALATWRAAGVTTPDAVGHSGQEIAAA VYAGALSLEDAAARVVVALRSRAITKLAGR
GMAAVELTADLEARLAPFGGRLAIATINSPHALVSGDDPAIDBELVELSQAOLFA
ASTPLAGRLRVAAGQAIRGVISFALDERPAPLEPSVPLGLASGLIQAALGDAID
APFLVTRGAVAGASERVTRFQAMTWGLGRVVALEHPRWGLVGLDGESEADANG
ARDANVHDRLANLVARDDQDLARAGTYVRLVRAPHDGAPFAPFAPFAPFAPFAPF
GTGALGHVARLARI GAHIVLAGRGAAPGAALAEEL SALGAVRFLAACDVADR
EAVALLPDRVEAGGTVPFVHAGGMAEPVAAWVELEADAIAAKAAGAOHLHDVF
AQRDLDFVLFSGAGVGGCQQAAYANAFLDAEARRADGLAATSTAWGAWAG
GMLTDAERELKRGVAPMDPVAALAHADHETCLTADVDWARFAPSASARP
RELLDELABARSALDAREPDDTAAPEPASTLRTTLAIPERHRLHUALVRT
ETAAVLGHADASRVENRFFDGLDLSMVELRRVRORATGKIPATLAFDHPETSA
LASKVLAIVLHDATPRASPAALERLEQLSAI VADEALRDLDLTARLAFDLKRAVR
TERPDDAAEAEKLGSADELIRLDIQKLGDRDIDVDY"
7542..22595
/gene="spH"
7542..22595
/gene="spH"
/function="biosynthesis of spirangiene"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="polyketide synthase"

```

```

/protein_id="CAD43449.1"
/db_xref="GI:27528335"
/db_xref="GOA:Q8GBX6"
/db_xref="SPTREMBL:Q8GBX6"
/translation="MSTVNDTLTETLRLTQELHSETRLRATEERRHEPIAIVGLG
VLEPGIHDRDITLWTELEGRDAIPIPASRWADATYDPPDAVGKSVYDAAMLDR
VLFDAADFGIIPREAKYVDPORILLETSMQLEADAGIVPASLRQSKTGCVFGTAS
SLRQGLDADSEAVAFMGWISFAGRLAFTLGLQGPALS IDTACSSSLVHLACQ
VIAQECEDLAVGVQVMSPEVFLVLSRTRALASGRKTSFANADGVRGEGVVVL
AVERLDRAKGRPIILAVGSAVNHGTSGLITVNGPQAQVLRALDADKLVAD
VVEVCHGTGTSIDPIEVNLAANYGEGRPKDRPLFGALKNTNIGHEFASGLAVA
KQVSAHVIIEAATPEETTPAALPAVAVLLSGKTDEALRAQARLHAHLIAGRPD
ARLVDAISLATRTHEDRAA VAADRELLGALDALARAGEAGCSVVASAIPAGRV
VVFPGQSGQVGMARALLASSVVRDEITACERALLAPHVMSIGVAILRGDEATIL
GRVVQVPLFAVVAALWRSIGVTPDAVVGSHQSEIATAATVAGALSLEDAKVA
LDRALTKTAGRMAAELGARDTEARLAPFGDAIATAINSRPRITVAGDIDAIDA
LVRLDEAQTIFARKVRVDYASHGAHVAIEIRELLADIAGEPRAGVPLYSVYTGAKL
DGRLDAAHFRNLRTVRDEATRALHDDGHRFFVEVSPHVPVIALALDITLSIGD
AAVSSLRDEGLERFLLSAGELHARGHAI DWGAFFAPLIGARRVTLPTVAFORFW
LDAGASDDADAPGASAEETA FWAL ERGDVAALSDALHVA DSGRSALSLMPALS
AMRSRREQSTUGAWRYRVEWRPVYSAASRGDVAGTVLVYAPAGVSDLASALAATE
RGADVPLALEASDI TAHDASRLDAIRGPAIRGVITLPTDPTDAPDRSPVAPAGL
ARSLAVQALAEIATAIPISIVTRGAVIGRSDRIESPQVAMWGLGRVVSLEQPERW
GGLVPLPRAHADGALDATALERLVAALGSEGEDELARPTGLFARLARAPLGDAPA
QRTPEGTVLVTCGTGALGAHVARLARGAHLVLA SRGAAAPGAELRAELTEIG
ARYTIACDVADRAAIEELFAALPAEPLTAI FHTAGVIDDGLVLSITPERLETVLHG
KADAARTLHETRLDLAAFLVPSAAGVFGNAGQAAVANAFLDAAEERALGIR
ATSVAMGAMAGMTFDVAQGLRGFGFAWEPEIIVAAALASALDRDTHVTIASIDW
ARFAPYASARQPLVDLPPARRALDAQPSATHEHDETALVATLRALSGEARHLV
SLVLAQTAAVQLGHADRLSDLPQRFGLDLSLMAVELRLLOKATKVLPAITITDR
PSHRLGFLVDLPLCGAASPSTADPRAARADASDEPIAIVGVGLPLPGVDVADAL
FRLEGREDAPVTPASRWDVGVDPDPPEAKGTYVRHAAMLDRVDFDAGFFGISP
REARVDPQRHLLETETAWALEAATIVPASIRUSTTCGVFGAGSAYAVLQSNBDEA
AYAMGTGAFAGRLAFTLGLQGPALSIDTACSSSLVHLACQSKRQECEDLAA
GVQVMAPEVFLVLSRTRALPQGRSKTFSANADGFGGEGVVLALERLDRARANG
FVLAVRGVTAHNDGASGSIATFNGSSQKVLRAALDDARLAPTDVVDVCHGTGTL
GDPIEVQALAAVYEGRAPERPLRLGAVKTNIGHEFASGLAGVVKIVASIRGHTLPA
TITHNPNPHDNDALPVVUDAAEPWERSNADSPBRAGVSAGLSCTNNAVVEEAP
EPDAPKAARGALPLUGSGLPVLSAKSDAALRAQAALRDQLARTPDAALDVAAALAT
TRSQFDRRAALASDHGALVAALLEGVAAPAGTIVAKGAADKLALFTGQGAQRAA
MGRGLYDAFVFRDALDVAHSLDRDLRDLVLFAPKGSSELASLDRTETFPALF
ALEVALPRLVEAMGVTPDVLGHSVGBELAAHVAGVLSLADACTVLAARLMOALPA
RDGAMVTYHAEVLAALRGDRABEIAA NAPSSTVAGVDVAVLVAARVFAEARG
KATLYRYSHA FSHHMDPMIDA FVEAEGLT FHPRIPIVSNVTGLAADEIRSPDYW
VRHVSAVRADGINTLEADGVSSFFELGPHGVLSALGESLANAQREAVFVPAIRDOG
RADVDATAUSSLHVHGRVDWAFFGPDPTVLTPTAFQDRFWLDRPAGATRSU
EAPAPEDAFWRAVDAGDVGALGATINAGSEDHLSALATLTPALSAMRRDERSL
VDALRYVWKPLPTTAARDVAGTWLVLTAGAEDALARELERALTARGAEVITMPTIA
PAEADRVLA VRVOGALADAGELRGIVSFAALDETTLATHTALPAGLALTURLVQALG
DMGLEAPLWLVTRGAVSTGRDLASAAQSNMTWGLRVVGLHPEHMGGLVDVADGDT
IDARALDRVTLASRGEQDLALRPTGIFARRLVRAPLGDAPARTYTTAGTSLVGT
GTGALGARVAFRAGGAEHLVLSRRGEDFAPGALRAELGARGVTLAACDVADR
ASVAAALRLDAQDQDPVRAVHAGVQAALASTDLDGQVQVAAKVRGALHILHELL
GDRELDALVLPASGAGVWGSCQQAAGNALFDALAEVRAAGITATSTAWGAWGG
GMLADADADAEEVLRKGLVPMAPALAAIACALAGHGETAVTVASVDWARFAPAFPA
SARPRLLDLP EAQRALEA PSAPAGADGASADRALVEALRPLSEADRARHLVELVA
ETAUVLGHADASADPHKGFDFLGLDSLVAVELLRLKLRARTGVLEPATVTFEDHPSR
LATVREALPDADAPADATATYSSSAHAADIEPIATVGLGLRPGGVEDREGINSF
LEQQRDGVGPTPKDRWDADAIYDPPPEAKGYSVREALLDRVDFDAFFGFSIPREA
KQVDPQHRLLLETAWOLEEATIVPASIKDTSTGVFGVIGLSDYALMQGSAEETDAY
AMGTLSSFAQRLAFTLGLQGPALS IDTACSSSLVHLACQSKRQECEDLALAGVQ
VMTLAEPPVLLSRTRALADGRSKTFSADHGYGRGEGVVVLALERLDRLEAARGRVVL
AVTRGAVNHGASGSIATPNGTSQKVLRAALGDAGLAPADVDECHGTGALGDP
IEVQALAAVYEGGREAERPLLGLAVKSNIGHEAASGLAGVAKIVTALEHGALPATIH
SRPNRPHVDMLVVRVDAELWARDEETPRAGVSAGLSGTNNAHIIIEEAPRAR
RGAATDPPPPVPLVPSAKSERALRAQERLRAHVTGRPGDALVDVAASLATTRSHF
BHRAAI VARDI GALEDALAGLEGRS PACTVTVSTTGKIALFTTQGGQQAAMGRGL
YGASVFRDADFACSYLDRELEHPLRDVVPAPRGSEL AALLDETLYTQTSILELEVA
LFLVLEAMGVTPHVLGHSTIGELVAHVAGVLSLEDACTLVAAARLIMHALPRDDAM
VAHACESEVLDALDAMDGRASIAAVNAPSSTVAVAGDREAIVAAVRAFEALGRKT
SRLRVSHA FSHHLDPMLEAFRRVAGVRFPPRIPIVSNVTGARASNDIDCSADYW
RHARGTVRFADGLEALRADGVSTFLELGHAVLALGHETLSESAVAFPLALDRGRDD

```

VDTLTAALGALHARGHEHVDWDAFFAFPNAGRVALTPTAFQRRFWLMDPKAKSSRAAA  
 ADSAGRYPLAGTRLDLPDGSVAHTLIDIGPQAOPYLADHSVYGRIVVPGAFHLAVLLAV  
 ABEHWPDSPLERLDVEFRALTFEPESETVAVRVHLTPAGDGGKGFHATVATQEGAM  
 THATATLGAAPPSSALARHPRAHEPLAGDEPSRAVARLMAFPAENWEPKRWLR  
 QTHAGDRTGVTGLAARDGVPDTDDADLPGLVDNAFAVYGSVAVTGLPADNBSVNP  
 PLPFTSERLVWYGRATPCWADFVMRGEITPPADRYIGDITILMSREGASSEGASRGST  
 PIATIEGLAFRRAPADKFLPELARNLYALTWPQEPAA SRPPRWALLIGEDSLGIAT  
 GIAADYPOVDALRODALRGAGPCPDVAVCLATAGHGSDDTDETELVS SVHTAARALA  
 LINAMLSRLERLASSTLVLTORAVATCAGEDVLDLASAALWGLVRSQAQENRDRAILL  
 VLDGSDASERALPYAVAGGALDDAPQLARDGRRLAPRLARVAAPATAANATST  
 APLRDPGTVLVTGCTGTGLVARRLVERHGVVHLVLLSRGPDAPCATDLABEL  
 QARGSVLAAADAEPALARVLLAI PHDRPLTAVVHAAGTLDLDDGVLSMTFARLSA  
 RLRKPAVALDWGVAOTSGITKHLEQADRLRIARGGLRPLASDEALF DLALAHRAA  
 VLPAAFDADALRKSGDALSALLRGLVAGPRACRAGAGAGASALKLRLASLPLRHH  
 RMLLDVADAVATVIGVASPLDI EPCRPQLQELGLDSLMAIELRNRLAAATGLRLSATL  
 LEDHPTPEALAKLISKLHDERAAPTSTSAAEIDRLLEGALASAYANEARIDELTVMK  
 AILSKWMAAREPSPDDGLASRLDANDDELFRMIDHVRABEAT"

gene  
 22592..33133  
 /gene="spiI"  
 CDS  
 22592..33133  
 /gene="spiI"  
 /function="biosynthesis of spirangiene"  
 /codon\_start=1  
 /transl\_table=11  
 /product="polyketide synthase"  
 /protein\_id="CAD43450.1"  
 /db\_xref="GI:27528336"  
 /db\_xref="GOA:Q8GBX5"  
 /db\_xref="SPTREMBL:Q8GBX5"  
 /translation="MSAPPGTEAKURDYLKQLTLDLQRSQTRLHAEASKAYEPIAIVA  
 MSCRFPGARSPEDLMQLLEGDRDVGSGFPASRGWSVEPRVDPRADVTGKTPALGEGF  
 LEGAELFDRAFFGVSPREALEMDPQORLLLETTWEAIEARAEIDPASLQGSQTGVGVV  
 MYNVAASHHLGGGDASGVYVSASSASVASGRITAYTFGRGPTVTVDTACSSSLVAVH

Query Match 4.0%; Score 20; DB 1; Length 4800;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	187	CCGCGACGCTGATCTTCGA	206
Db	38478	CCGCGACGCTGATCTTCGA	38497

Search completed: June 16, 2004, 17:31:13  
 Job time : 2084.33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:31:50 ; Search time 246 Seconds  
(without alignments)  
8651.818 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcggtcccgacgcgcgagca.....ccctcgggcagagactga 501

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2587

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

1: geneseqn1980a:\*

2: geneseqn1990a:\*

3: geneseqn2000a:\*

4: geneseqn2001a:\*

5: geneseqn2001bs:\*

6: geneseqn2002a:\*

7: geneseqn2003a:\*

8: geneseqn2003bs:\*

9: geneseqn2003cs:\*

10: geneseqn2004a:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	53789	AAV21187	AAV21187 Amycolato
2	31	6.2	9222	AAU61171	AAU61171 Actinosyn
3	31	6.2	82746	AAU61224	AAU61224 Actinosyn
4	25	5.0	44377	AAT80414	AAT80414 Platenoli
5	25	5.0	44377	AAT78508	AAT78508 Platenoli
6	23	4.6	748	AAU89129	AAU89129 Polyketid
7	23	4.6	753	AAU89143	AAU89143 Polyketid
8	23	4.6	6378	ACF39397	ACF39397 Mycobacte
9	23	4.6	31422	AAU92302	AAU92302 S. avermi
10	23	4.6	31422	AAU79278	AAU79278 Streptomy
11	23	4.6	110000	AAI99682	AAI99682 18 o
12	23	4.6	110000	AAI99683	AAI99683 18 o
13	21	4.2	47981	AAU30757	AAU30757 Micromono
14	20	4.0	4551	ACA42620	ACA42620 Prokaryot
15	20	4.0	29879	AAU46806	AAU46806 eryA regi
16	20	4.0	31422	AAU92302	AAU92302 S. avermi
17	20	4.0	31422	AAU79278	AAU79278 Streptomy
18	20	4.0	43280	AAT80413	AAT80413 Ty lactone
19	19	3.8	177	AAU7049	AAU7049 Human bre
20	19	3.8	177	AAU08770	AAU08770 Human bre
21	19	3.8	375	AAT73432	AAT73432 Human pre
22	19	3.8	375	ACA63335	ACA63335 DNA encod
23	19	3.8	375	AAZ35707	AAZ35707 Human uro

24	19	3.8	452	2	AAV99120	AAV99120 DNA methy
25	19	3.8	953	6	ABS63429	ABS63429 DNA encod
26	19	3.8	1049	6	ABS63428	ABS63428 DNA encod
27	19	3.8	1368	2	AAQ82825	AAQ82825 CDNA enco
28	19	3.8	1698	2	AAU86613	AAU86613 CDNA enco
29	19	3.8	1893	4	AAU54382	AAU54382 Pseudomon
30	19	3.8	1893	7	ACA42802	ACA42802 Prokaryot
31	19	3.8	2040	6	ABN85313	ABN85313 Human cyt
32	19	3.8	2448	9	ADD29815	ADD29815 Human tum
33	19	3.8	2478	2	AAU86614	AAU86614 CDNA enco
34	19	3.8	3081	6	ABQ88107	ABQ88107 Human oet
35	19	3.8	5092	4	AAI59458	AAI59458 Human pol
36	19	3.8	5434	6	ABL62902	ABL62902 Breast ca
37	19	3.8	5434	6	ABL63111	ABL63111 Breast ca
38	19	3.8	5434	6	ABL64407	ABL64407 Stomach c
39	19	3.8	5434	7	ABT33842	ABT33842 DNMT1 cDN
40	19	3.8	5794	5	AAU78583	AAU78583 DNA encod
41	19	3.8	19440	2	AAU99129	AAU99129 DNA methy
42	19	3.8	113193	7	AAU54645	AAU54645 Streptomy
43	19	3.8	128139	6	AAI64291	AAI64291 RRV genom
44	19	3.8	133719	3	AAU64754	AAU64754 Macaca mu
45	18	3.6	276	7	ACD93990	ACD93990 Human col

## ALIGNMENTS

## RESULT 1

AAV21187

ID AAV21187 standard; DNA; 53789 BP.

XX AAV21187;

XX 24-JUL-1998 (first entry)

XX Amycolatopsis mediterranei

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

KW polyketide synthase; actinomycete; ansamycin; ds.

XX Amycolatopsis mediterranei.

XX Key Location/Qualifiers

FT CDS 1825..15543

FT /tag= a

FT /label= ORF\_A

FT /product= "Polyketide synthase"

FT 15550..30759

FT /tag= b

FT /label= ORF\_B

FT /product= "polyketide synthase"

FT 30895..36060

FT /tag= C

FT /label= ORF\_C

FT /product= "Polyketide synthase"

FT 36259..41325

FT /tag= d

FT /label= ORF\_D

FT /product= "Polyketide synthase"

FT 41373..51614

FT /tag= e

FT /label= ORF\_E

FT /product= "polyketide synthase"

FT 51713..5293

FT /tag= f

FT /label= ORF\_F

FT /product= "polyketide synthase"

XX WO9807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997;

XX 97WO-EP004495.

XX 20-AUG-1996; 96EP-00810551.  
XX (NOVS ) NOVARTIS AG.  
XX Schupp T, .Toupet C, Engel N;  
XX WPI; 1998-169172/15.  
XX P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.  
XX  
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used to  
XX produce rifamycin and rifamycin analogues.  
XX  
XX Claim 4; Page 53-102; 205pp; English.  
XX  
XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
XX synthesis gene cluster DNA fragment from the present invention. The DNA  
XX fragment comprises a DNA region involved directly or indirectly in the  
XX gene cluster responsible for rifamycin synthesis, including the adjacent  
XX DNA regions to the right and left which, by reason of their function in  
XX connection with rifamycin biosynthesis, qualify as constituents of this  
XX rifamycin gene cluster, and functional fragments, derivatives or  
XX constituents of these. The Amycolatopsis mediterranei rifamycin synthesis  
XX gene cluster DNA fragment can be used for producing rifamycin, rifamycin  
XX analogues or precursors. It can also be used for inactivating or  
XX modifying genes involved in ansamycin or rifamycin biosynthesis. The DNA  
XX can be used for constructing mutant actinomycetes strains from which the  
XX natural rifamycin or ansamycin biosynthesis gene cluster has been partly  
XX or completely deleted. The DNA fragment can be used for assembling a  
XX library of polyketide synthases, which can be used for assembling a  
XX library of polyketides. A hybridisation probe of the invention can be  
XX used for identifying DNA fragments involved in the biosynthesis of  
XX ansamycins  
XX  
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 U; 0 Other;  
Query Match 100.0%; Score 501; DB 2; Length 53789;  
Best Local Similarity 100.0%; Pred. No. 1.7e-218;  
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTCTCCGAGCGGAGCAACCGATCTGCTGAAGCTGTTCGGGGGACGCTTCG 60  
DB 15043 GCGTCTCCGAGCGGAGCAACCGATCTGCTGAAGCTGTTCGGGGGACGCTTCG 15102  
QY 61 ACGETGTCTGCGCCACAGCGGCGCGAGGCGATCGCCCGCCAGGCGTTCAGGAGTGC 120  
DB 15103 ACGGTGCTTCGGCCACAGCGGCGCGAGGCGATCGCCCGCCAGGCGTTCAGGAGTGC 15162  
QY 121 GGCTTCGACTCGTGGCGCGGCTCAACCTCGCAACAGCCTGCAACCGCGCCACCGGGCTG 180  
DB 15163 GGCCTTCGACTCGTGGCGCGGCTCAACCTCGCAACAGCCTGCAACCGCGCCACCGGGCTG 15222  
QY 181 CGGTGCGCGGAGCGCTGATCTTCACTACCCACCGGAGGCGTGTGGGTACCTG 240  
DB 15223 CGGTGCGCGGAGCGCTGATCTTCACTACCCACCGGAGGCGTGTGGGTACCTG 15282  
QY 241 CGCGTCGAACCTCTCGGAGGCGCGAGCGACGCGCTGGAAGCGGCGAAGACCTCCGG 300  
DB 15283 CGCGTCGAACCTCTCGGAGGCGCGAGCGACGCGCTGGAAGCGGCGAAGACCTCCGG 15342  
QY 301 CGAGTCTCTCGCGCGCGTGCCTGCTCGCCGGTTCAAGAGGCGCGGCGTGTGACACGCTG 360  
DB 15343 CGAGTCTCTCGCGCGCGTGCCTGCTCGCCGGTTCAAGAGGCGCGGCGTGTGACACGCTG 15402  
QY 361 CTCGGCTTCGCGACACCGGCGCGAGACCGGCGCGAGACCGGCGCGAGCGCGCC 420  
DB 15403 CTCGGCTTCGCGACACCGGCGCGAGACCGGCGCGAGACCGGCGCGAGCGCGCC 15462  
QY 421 CCGGGCGCGGAGCGCAACCTGATCGACGCGCTGACATCTCGGGTCTCGTCAACGA 480  
DB 15463 CCGGGCGCGGAGCGCAACCTGATCGACGCGCTGACATCTCGGGTCTCGTCAACGA 15522  
QY 481 GGCCTCGGGGAGACGAGCTGA 501

DB 15523 GCCCTCGGGCAGCAGCTGA 15543  
RESULT 2  
AAL61171  
ID AAL61171 standard; DNA; 9222 BP.  
XX  
XX AAL61171;  
XX  
XX 22-SEP-2003 (first entry)  
XX  
XX Actinosynnema pretiosum polyketide synthase (PKS) gene #2.  
XX  
XX Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
XX gene; ds.  
XX  
XX Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.  
XX  
XX 21-NOV-2002; 2002WO-US037547.  
XX  
XX 21-NOV-2001; 2001US-0332158P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Floss HG, Yu T, Leistner E;  
XX  
XX WPI; 2003-493374/46.  
XX  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Disclosure; Page 53-58; 160pp; English.  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumor agent. The present sequence is A. pretiosum  
XX ansamitocin gene cluster I polyketide synthase (PKS) gene  
XX  
XX Sequence 9222 BP; 784 A; 3437 C; 3858 G; 1143 T; 0 U; 0 Other;  
Query Match 6.2%; Score 31; DB 7; Length 9222;  
Best Local Similarity 100.0%; Pred. No. 0.00017;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 168 GGCACACGGGCTGGGCTGCCCGGAGCGCTG 198  
DB 5130 GGCACACGGGCTGGGCTGCCCGGAGCGCTG 5160  
RESULT 3  
AAL61224/c  
ID AAL61224 standard; DNA; 82746 BP.  
XX  
XX AAL61224;  
XX  
XX 22-SEP-2003 (first entry)  
XX  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX  
XX Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.

```
XX 21-NOV-2002; 2002WO-US037547.
XX PF
XX 21-NOV-2001; 2001US-0332158P.
XX PR
XX (UNIW ) UNIV WASHINGTON.
XX PA
XX Floss HG, Yu T, Leistner E;
XX PI
XX WPI; 2003-493374/46.
XX DR
XX Novel maytansinoid produced by bacterial host cell transformed with
XX PT expression vector comprising open reading frame from ansamitocin gene
XX PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX PS
XX Claim 7; Page 105-152; 160pp; English.
XX CC
XX The invention relates to maytansinoid produced by bacterial host cell
XX CC transformed with expression vector comprising open reading frame from
XX CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX CC useful as a potent antitumor agent. The present sequence is A. pretiosum
XX CC ansamitocin ansamitocin biosynthetic gene cluster I
XX CC
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;
SQ
Query Match 6.2%; Score 31; DB 7; Length 82746;
Best Local Similarity 100.0%; Pred.No. 0.00013;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 168 GGCACCGGGCTGGCTGGCGGCGGACGCTG 198
Db 17833 GGCACCGGGCTGGCTGGCGGCGGACGCTG 17803
RESULT 4
AAT80414
ID AAT80414 standard; DNA; 44377 BP.
XX AC
XX AAT80414;
XX DT 27-FEB-1998 (first entry)
XX DE
XX Platenolide synthase gene cluster.
XX KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
XX KW platenolide synthase gene cluster; platenolide production; smg gene;
XX KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
XX OS Streptomyces ambofaciens.
XX FH
XX Key Location/Qualifiers
XX CDS 350..14002
XX FT /*tag= a
XX FT /transl_except= (pos:350..352, aa:Met)
XX FT /note= "ORF1 encodes protein shown in AAW22606"
XX CDS 14046..20036
XX FT /*tag= b
XX FT /note= "ORF2 encodes protein shown in AAW22607"
XX CDS 20110..31284
XX FT /*tag= c
XX FT /transl_except= (pos:20111..20113, aa:Met)
XX FT /note= "ORF3 encodes protein shown in AAW22608"
XX CDS 31329..36071
XX FT /*tag= d
XX FT /note= "ORF4 encodes protein shown in AAW22609"
XX CDS 36155..41830
XX FT /*tag= e
XX FT /note= "ORF5 encodes protein shown in AAW22610"
XX PN EP791655-A2.
XX XX
XX 27-AUG-1997.
XX PD
XX
```

```
PF 19-FEB-1997; 97EP-00301056.
XX XX
PR 22-FEB-1996; 96US-0012078P.
XX XX
XX (ELIL ) LILLY & CO ELI.
XX PA
XX Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX PI
XX WPI; 1997-418046/39.
XX DR
XX P-PSDB; AAW22606, AAW22607, AAW22608, AAW22609, AAW22610.
XX DR
XX DNA encoding Streptomyces fradiae ty lactone synthase domain - for
XX PT production of tylosin-related polyketide compounds.
XX PT
XX Example 2; Page 110-134; 220pp; English.
XX PS
XX This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the smg gene, and was
XX CC isolated from Streptomyces ambofaciens. This sequence encodes the multi-
XX CC functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macrolide
XX CC antibiotic spiramycin. This sequence was used along with the tylG gene
XX CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
XX CC gene is the ty lactone synthase gene cluster of the invention. The tylG
XX CC sequence was isolated from Streptomyces fradiae, and encodes
XX CC multifunctional proteins which direct the synthesis of the polyketide
XX CC ty lactone. Ty lactone is the basic building block of the antibiotic
XX CC tylosin. The hybrid sequence can be used to transform S. ambofaciens
XX CC lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1
XX CC sequence, so that they can produce polyketides. The DNA sequence can be
XX CC modified so as to alter the type of carboxylic acids incorporated, the
XX CC number of carboxylic acids incorporated and/or the post-condensation
XX CC reactions performed, thereby resulting in novel tylosin-related
XX CC polyketides
XX CC
XX Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;
SQ
Query Match 5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred.No. 0.076;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 119 TCGGCTTCGACTCGCTGGCGCGGT 143
Db 35710 TCGGCTTCGACTCGCTGGCGCGGT 35734
RESULT 5
AAT78508
ID AAT78508 standard; DNA; 44377 BP.
XX AC
XX AAT78508;
XX DT 26-FEB-1998 (first entry)
XX DE
XX Platenolide synthase gene cluster.
XX KW Platenolide synthase gene cluster; platenolide production; smg gene;
XX KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX OS Streptomyces ambofaciens.
XX FH
XX Key Location/Qualifiers
XX CDS 350..14002
XX FT /*tag= a
XX FT /transl_except= (pos:350..352, aa:Met)
XX FT /note= "ORF1 encodes protein shown in AAW23716"
XX CDS 14046..20036
XX FT /*tag= b
XX FT /note= "ORF2 encodes protein shown in AAW23717"
XX CDS 20110..31284
XX FT /*tag= c
XX FT /transl_except= (pos:20111..20113, aa:Met)
XX FT /note= "ORF3 encodes protein shown in AAW23718"
XX PD
```

```
FT CDS 31329..36071
FT /*tag= d
FT /note= "ORF4 encodes protein shown in AAW23719"
FT CDS 36155..41830
FT /*tag= e
FT /note= "ORF5 encodes protein shown in AAW23720"
FT XX
FT EP791656-A2.
FT XX
FT 27-AUG-1997.
FT XX
FT 19-FEB-1997; 97EP-00301066.
FT XX
FT 22-FEB-1996; 96US-0012050P.
FT XX
FT (ELIL ) LILLY & CO ELI.
FT XX
FT Burgett SG, Kuhstoss SA, Rao RN, Richardson MA, Rosteck PR;
FT PI
FT XX
FT WPI; 1997-418047/39.
FT DR
FT P-PSDB; AAW23716, AAW23717, AAW23718, AAW23719, AAW23720.
FT DR
FT DNA encoding Streptomyces ambofaciens platenolide synthase domain - for
FT production of epiramycin-related polyketide antibiotics.
FT PT
FT Claim 9; Page 8-33; 81pp; English.
FT PS
FT XX
FT This sequence represents the platenolide synthase gene cluster of the
FT invention. This sequence is referred to as the smg gene, and was
FT isolated from Streptomyces ambofaciens. This sequence encodes the multi-
FT functional proteins which direct the synthesis of the polyketide
FT platenolide. Platenolide is the basic building block of the macrolide
FT antibiotic spiramycin. The DNA can be used to produce compounds
FT exhibiting antibiotic activity based on the platenolide structure.
FT including specifically the macrolide antibiotic spiramycin and spiramycin
FT analogues and derivatives. Modifications of the platenolide synthase DNA
FT sequence can be made so as to change the number and type of carboxylic
FT acids incorporated into the growing polyketide chain and to change the
FT kind of post-condensation processing that is conducted
FT XX
FT SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;
FT
FT Query Match 5.0%; Score 25; DB 2; Length 44377;
FT Best Local Similarity 100.0%; Pred. No. 0.076;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT Qy 119 TCGGCTTCGACTCGTGGCCGCGGT 143
FT Db 35710 TCGGCTTCGACTCGTGGCCGCGGT 35734
FT
FT RESULT 6
FT AAW89129/c
FT ID AAW89129 standard; DNA; 748 BP.
FT XX
FT AC AAW89129;
FT AC
FT DT 19-MAR-2001 (first entry)
FT XX
FT DE Polyketide synthase derived sequence 34-183R1A1.T7 DNA.
FT XX
FT KW Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
FT KW immunosuppressive; neuroimmunophilin; ds.
FT XX
FT OS Streptomyces hygroscopicus.
FT XX
FT Key Location/Qualifiers
FT FT 1..747
FT FT /*tag= a
FT FT /partial
FT XX
FT US6150513-A.
FT PN
FT XX
```

```
PD 21-NOV-2000.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX 16-SEP-1998; 98US-00154083.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Wu K;
XX
XX WPI; 2001-049104/06.
XX P-PSDB; AAB19896.
XX
XX Novel nucleic acid encoding polyketide synthase useful for producing
XX recombinant polyketide synthase for making novel polyketides for use as
XX antibiotic, immunosuppressant or as neuroimmunophilin.
XX
XX Example; Col 20-21; 30pp; English.
XX
XX The present sequence is that of a polyketide synthase (PKS)-derived
XX coding sequence designated 34-183R1A1.T7. The DNA was isolated from
XX cosmid 34-183 of an FK-520-producing strain of Streptomyces hygroscopicus
XX following restriction digestion. It does not comprise a full-length
XX coding sequence. The 249-amino acid translated sequence is given in
XX AAB19896. An isolated nucleic acid comprising a coding sequence for a PKS
XX enzyme identical to, or isolated from, a PKS coding sequence contained
XX within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is
XX claimed. Such nucleic acids, when used alone or in combination with other
XX PKS domain coding sequences, are useful in the construction of
XX recombinant vectors that encode PKS enzymes which can be used to make
XX novel polyketides having applications in medicine, agriculture and animal
XX health. The polyketides may have e.g. antibiotic, immunosuppressive or
XX neuroimmunophilin activities
XX
XX SQ Sequence 748 BP; 114 A; 250 C; 277 G; 107 T; 0 U; 0 Other;
XX
XX Query Match 4.6%; Score 23; DB 4; Length 748;
XX Best Local Similarity 100.0%; Pred. No. 1;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 169 GCCACCGGCTGGCGTCCCGC 191
XX Db 366 GCCACCGGCTGGCGTCCCGC 344
XX
XX RESULT 7
XX AAW89143/c
XX ID AAW89143 standard; DNA; 753 BP.
XX XX
XX AC AAW89143;
XX AC
XX DT 19-MAR-2001 (first entry)
XX XX
XX DE Polyketide synthase derived sequence 34-183R1A.2 DNA.
XX XX
XX KW Polyketide synthase; PKS; cosmid 34-183; FK-520; antibiotic;
XX KW immunosuppressive; neuroimmunophilin; ds.
XX XX
XX OS Streptomyces hygroscopicus.
XX XX
XX FN US6150513-A.
XX XX
XX PD 21-NOV-2000.
XX XX
XX PF 16-SEP-1998; 98US-00154083.
XX XX
XX PR 16-SEP-1998; 98US-00154083.
XX XX
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX FI Wu K;
XX XX
XX WPI; 2001-049104/06.
XX
```

XX Novel nucleic acid encoding polyketide synthase useful for producing  
PT recombinant polyketide synthase for making novel polyketides for use as  
PT antibiotic, immunosuppressant or as neuroimmunophilin.  
XX Example; Col 27; 30pp; English.  
XX  
XX The present sequence is that of a DNA sequence designated 34-183RIA.2,  
CC which was isolated from cosmid 34-183 of an FX-520-producing strain of  
CC Streptomyces hygroscopicus following restriction digestion. An isolated  
CC nucleic acid comprising a coding sequence for a polyketide synthase (PKS)  
CC enzyme identical to, or isolated from, a PKS coding sequence contained  
CC within a cosmid selected from cosmids 34-183, 34-122 and 34-126 is  
CC claimed. Such nucleic acids, when used alone or in combination with other  
CC PKS domain coding sequences, are useful in the construction of  
CC recombinant vectors that encode PKS enzymes which can be used to make  
CC novel polyketides having applications in medicine, agriculture and animal  
CC health. The polyketides may have e.g. antibiotic, immunosuppressive or  
CC neuroimmunophilin activities  
XX  
XX Sequence 753 BP; 115 A; 253 C; 277 G; 108 T; 0 U; 0 Other;  
SQ  
Query Match 4.6%; Score 23; DB 4; Length 753;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 169 GCCACCGGCTGGCGCTGCCGCGC 191  
DB 371 GCCACCGGCTGGCGCTGCCGCGC 349  
RESULT 8  
ACF39397  
XX ACF39397  
XX  
XX 26-SEP-2003 (first entry)  
XX  
XX Mycobacterium tuberculosis mycobacterial antigen DNA SEQ ID NO:172.  
DE  
XX  
XX Mycobacterium tuberculosis; mycobacterial; antigen; infection; vaccine;  
XX tuberculostatic; mycobacterial peptide; mycobacterial infection; gene;  
XX ds.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO2003033530-A2.  
XX  
XX 24-APR-2003.  
XX  
XX 14-OCT-2002; 2002WO-GB004647.  
XX  
XX 12-OCT-2001; 2001GB-00024593.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX James B, Bacon J, March P;  
XX  
XX WPI; 2003-393501/37.  
XX P-PSDB; ABM15903.  
XX  
XX New isolated mycobacterial peptide encoded by a gene that is induced or  
PT up-regulated under high oxygen tension, useful for diagnosing, treating  
PT or preventing a mycobacterial infection.  
XX  
XX Claim 18; Page 302-306; 392pp; English.  
XX  
XX The present invention describes an isolated mycobacterial peptide (I), or  
CC its fragment, variant or derivative encoded by a gene whose expression is  
CC induced or up-regulated during culture of a mycobacterium under  
CC continuous culture conditions of a dissolved oxygen tension of at least  
CC 30% air saturation measured at 37 plus degrees Celsius when compared with

CC a dissolved oxygen tension of up to 10% air saturation measured at 37  
CC plus degrees Celsius. (I) has tuberculostatic activity and can be used in  
CC vaccines. The mycobacterial peptide (I) or its fragment, variant or  
CC derivative, inhibitor, antibody, attenuated mycobacterium, attenuated  
CC microbial carrier, DNA sequence, DNA plasmid, RNA sequence, or RNA vector  
CC from the present invention can be used for manufacturing a medicament for  
CC treating or preventing a mycobacterial infection. The peptide or its  
CC fragment, variant or derivative, the antibody, or a polynucleotide probe  
CC comprising at least 8 nucleotides, where the probe binds to at least a  
CC part of the gene, is useful for manufacturing a diagnostic reagent for  
CC identifying a mycobacterial infection. The present sequence encodes a  
CC Mycobacterium tuberculosis mycobacterial antigen, which is used in the  
CC exemplification of the present invention  
XX  
XX Sequence 6378 BP; 815 A; 1685 C; 2614 G; 1264 T; 0 U; 0 Other;  
SQ

Query Match 4.6%; Score 23; DB 7; Length 6378;  
Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGCTGATCTTCGACTACCCAC 215  
DB 6094 ACGCTGATCTTCGACTACCCAC 6116

RESULT 9

AAA92302

ID AAA92302 standard; DNA; 31422 BP.

XX AAA92302;

XX 10-JAN-2001 (first entry)

DE S. avermitilis avermectin aglycon synthase DNA aveAll SEQ ID NO:2.

XX Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
XX multifunctional enzyme; polyketide; avermectin; veterinary drug;  
XX agrochemical; ds.  
XX  
XX Streptomyces avermitilis.  
XX  
XX Key Location/Qualifiers  
XX CDS 1. 14646  
XX FT /tag= a  
XX FT /note= "avermectin aglycon synthase protein"  
XX FT 14824. 31422  
XX FT /tag= b  
XX FT /note= "avermectin aglycon synthase protein"  
XX  
XX WO200050605-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX 23-FEB-2000; 2000WO-JP001041.  
XX  
XX 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
XX Omura S, Ikeda H;  
XX  
XX WPI; 2000-565458/52.  
XX P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
XX Claim 2; Page 134-203; 314pp; Japanese.  
XX  
XX The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host

CC cells transformed by the vectors; (4) preparation of the polypeptides by  
 CC culture of the transformants; (5) preparation of avermectin aglycon or  
 CC its derivatives by culture of transformed avermectin-producing  
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
 CC containing sense or antisense sequences from the avermectin aglycon  
 CC synthase DNA. The enzymes are useful for the production of modified forms  
 CC of avermectin and of the intermediates in its biosynthesis, for use as  
 CC drugs, veterinary drugs and agrochemicals  
 XX  
 SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;

Query Match 4.6%; Score 23; DB 3; Length 31422;  
 Best Local Similarity 100.0%; Pred. No. 0.64;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 193 AGCGTGATCTTCGACTACCCAC 215  
 DB 4645 AGCGTGATCTTCGACTACCCAC 4667

RESULT 10

AAH79278  
 ID AAH79278 standard; DNA; 31422 BP.

XX AC AAH79278;

XX 04-DEC-2001 (first entry)

XX Streptomyces avermitilis coding sequences SEQ ID NO: 2.

XX Avermectin aglycone synthase; AAS; avermectin derivative;  
 KW drug production; veterinary drug; pesticide; ds.

XX Streptomyces avermitilis.

XX Key Location/Qualifiers

FT CDS 1. .14646

FT /\*tag= a

FT /product= "AAG65266"

FT 14824. .31422

FT /\*tag= b

FT /product= "AAG65267"

XX WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP001381.

XX 24-FEB-2000; 2000JP-00047405.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX (KITA ) KITASATO INST.

XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX WPI; 2001-582053/65.

XX P-FSB; AAG65266, AAG65267.

XX New modified avermectin aglycone synthase derived from Streptomyces  
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
 PT drugs and pesticides.

XX Disclosure; Page 103-149; 257pp; Japanese.

XX The present invention relates to the production of modified derivatives  
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
 CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
 CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
 CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
 CC can be used in the production of drugs, veterinary drugs and pesticides.  
 CC The present sequence is a fragment of the S. avermitilis genome

XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;  
 Query Match 4.6%; Score 23; DB 4; Length 31422;  
 Best Local Similarity 100.0%; Pred. No. 0.64;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 193 AGCGTGATCTTCGACTACCCAC 215  
 DB 4645 AGCGTGATCTTCGACTACCCAC 4667

RESULT 11

AAI99682\_18

Continuation (19 of 45) of AAI99682 from base 1800001 (Mycobacterium tuberculosis strai

WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 4.6%; Score 23; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0.55;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 AGCGTGATCTTCGACTACCCAC 215

DB 81395 AGCGTGATCTTCGACTACCCAC 81417

RESULT 12





```
FT /function= "KS1"
FT 15427. .16476
FT /*tag= p
FT /gene= "megA"
FT /function= "AT1"
FT 17155. .17694
FT /*tag= q
FT /gene= "megA"
FT /function= "KR1"
FT 17947. .18207
FT /*tag= r
FT /gene= "megA"
FT /function= "ACP1"
FT 18268. .19548
FT /*tag= s
FT /gene= "megA"
FT /function= "KS2"
FT 19876. .20910
FT /*tag= t
FT /gene= "megA"
FT /function= "AT2"
FT 21517. .22053
FT /*tag= u
FT /gene= "megA"
FT /function= "KR2"
FT 22318. .22575
FT /*tag= v
FT /gene= "megA"
FT /function= "ACP2"
FT 22867. .33555
FT /*tag= w
FT /gene= "megA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT /note= "polyketide synthase, encodes AAB82213"
FT 22957. .24237
FT /*tag= x
FT /gene= "megAII"
FT /function= "KS3"
FT 22957. .24237
FT /*tag= ai
FT /gene= "megAII"
FT /function= "KS5"
FT 24544. .25581
FT /*tag= y
FT /gene= "megAII"
FT /function= "AT3"
FT 24544. .25581
FT /*tag= aj
FT /gene= "megAII"
FT /function= "AT5"
FT 26230. .26733
FT /*tag= z
FT /gene= "megAII"
FT /function= "KR3 (inactive)"
FT 26230. .26733
FT /*tag= ak
FT /gene= "megAII"
FT /function= "KR5"
FT 26998. .27258
FT /*tag= aa
FT /gene= "megAII"
FT /function= "ACP3"
FT 26998. .27258
FT /*tag= al
FT /gene= "megAII"
FT /function= "ACP5"
FT 27393. .28590
FT /*tag= ab
FT /gene= "megAII"
FT /function= "KS4"
FT 27393. .28590
FT /*tag= am
FT /gene= "megAII"
```

```
FT misc_feature /function= "KS6"
FT 28897. .29931
FT /*tag= ac
FT /gene= "megAII"
FT /function= "AT4"
FT 28897. .29931
FT /*tag= an
FT /gene= "megAII"
FT /function= "AT6"
FT 29953. .30477
FT /*tag= ad
FT /gene= "megAII"
FT /function= "DH4"
FT 29953. .30477
FT /*tag= ao
FT /gene= "megAII"
FT /function= "KR6"
FT 31396. .32244
FT /*tag= ae
FT /gene= "megAII"
FT /function= "ER4"
FT 31396. .32244
FT /*tag= ap
FT /gene= "megAII"
FT /function= "ACP6"
FT 32257. .32799
FT /*tag= af
FT /gene= "megAII"
FT /function= "KR4"
FT 32257. .32799
FT /*tag= aq
FT /gene= "megAII"
FT /function= "TE"
FT 33052. .33312
FT /*tag= ag
FT /gene= "megAII"
FT /function= "ACP4"
FT 33666. .43271
FT /*tag= ah
FT /gene= "megAII"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 3"

Query Match 4.2%; Score 21; DB 4; Length 47981;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTCGACTCGCTGGCGGGGTC 144
DB 22447 TTCGACTCGCTGGCGGGGTC 22467

RESULT 14
ACA42620
ID ACA42620 standard; DNA; 4551 BP.
XX
AC ACA42620;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #24277.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas aeruginosa.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
```

```
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-023926/02.
XX P-PSDB; ABU38750.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 30490; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 4551 BP; 779 A; 1608 C; 1519 G; 645 T; 0 U; 0 Other;
XX
XX Query Match 4.0%; Score 20; DB 7; Length 4551;
XX Best Local Similarity 100.0%; Pred. No. 19;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 258 GGAGCCCGACGACGCGCTGG 277
XX |||||
XX Db 3600 GGAGCCCGACGACGCGCTGG 3619
XX
XX RESULT 15
XX AAQ46806
XX ID AAQ46806 standard; DNA; 29879 BP.
XX
XX AC AAQ46806;
XX
XX XX 27-AUG-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 22-DEC-1993 (first entry)
XX
XX eryA region of S. erythraea chromosome.
XX
```

```
XX Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
XX erythromycin; condensation; elongation; acyl chain growth;
XX gene replacement; ss.
XX
XX Saccharopolyspora erythraea.
XX
XX Key Location/Qualifiers
XX CDS 3..10199
XX /*tag= a
XX /note= "ORF 1"
XX 10218..20921
XX /*tag= b
XX /note= "ORF 2"
XX 20922..29879
XX /*tag= C
XX /note= "ORF 3"
XX
XX WO9313663-A1.
XX
XX 22-JUL-1993.
XX
XX 17-JAN-1992; 92WO-US000427.
XX
XX 17-JAN-1992; 92WO-US000427.
XX (ABBO ) ABBOTT LAB.
XX
XX Katz L, Donadio S, Mcalpine JB;
XX WPI; 1993-242804/30.
XX P-PSDB; AAR44430, AAR44431, AAR44432.
XX
XX Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -
XX by introducing altered biosynthetic gene-contg. DNA into microorganisms.
XX
XX Claim 27; Fig 2; 133pp; English.
XX
XX This sequence represents a fragment of the Saccharopolyspora erythraea
XX genome, designated eryA. The polypeptides encoded by this region are
XX involved in the biosynthesis of the polyketide segment of erythromycin.
XX eryA is organised in modules and each module takes care of one
XX condensation step. The precise succession of elongation steps is dictated
XX by the genetic order of the modules. This fragment may be specifically
XX altered such that novel polyketide molecules of desired structure are
XX produced. Three types of alteration may be produced; those inactivating a
XX single function in a module which does not arrest acyl chain growth;
XX those inactivating a single function in a module which does affect chain
XX growth; and those affecting an entire module. The mutations may be
XX introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN
XX field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;
XX
XX Query Match 4.0%; Score 20; DB 2; Length 29879;
XX Best Local Similarity 100.0%; Pred. No. 15;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 124 TTCGACTCGTGGCGCGGT 143
XX |||||
XX Db 20553 TTCGACTCGTGGCGCGGT 20572
XX
XX RESULT 16
XX AA92302/c
XX ID AA92302 standard; DNA; 31422 BP.
XX
XX AC AA92302;
XX
XX XX 10-JAN-2001 (first entry)
XX DT
XX DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
```

KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical; db.  
OS Streptomyces avermitilis.  
XX  
XX Key Location/Qualifiers  
FH 1. .14646  
FT /\*tag= a  
FT /product= "AAG65266"  
FT 14824. .31422  
FT /\*tag= b  
FT /product= "AAG65267"  
XX WO200050605-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX /note= "avermectin aglycon synthase protein"  
FT 14824. .31422  
FT /\*tag= b  
FT /note= "avermectin aglycon synthase protein"  
XX  
XX WO200050605-A1.  
XX  
XX 31-AUG-2000.  
XX  
XX 23-FEB-2000; 2000WO-JP001041.  
XX  
XX 24-FEB-1999; 99JP-00046961.  
XX  
XX (KITA ) KITASATO INST.  
XX  
XX Omura S, Ikeda H;  
XX  
XX WPI; 2000-565458/52.  
DR P-PSDB; AAB23751, AAB23752.  
XX  
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of  
PT it for the production of avermectin and its derivatives for drug and  
PT agrochemical use.  
XX  
XX Claim 2; Page 134-203; 314pp; Japanese.  
XX  
XX The present sequence represents DNA which encodes avermectin aglycon  
CC synthase proteins. Also described are: (1) polypeptides encoded by all or  
CC part of the DNA; (2) expression vectors containing the DNA; (3) host  
CC cells transformed by the vectors; (4) preparation of the polypeptides by  
CC culture of the transformants; (5) preparation of avermectin aglycon or  
CC its derivatives by culture of transformed avermectin-producing  
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length  
CC containing sense or antisense sequences from the avermectin aglycon  
CC synthase DNA. The enzymes are useful for the production of modified forms  
CC of avermectin and of the intermediates in its biosynthesis, for use as  
CC drugs, veterinary drugs and agrochemicals  
XX  
XX Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 U; 0 Other;  
SQ Query Match 4.0%; Score 20; DB 3; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 366 CCTCGCGACACCGGCACCG 385  
Db 27210 CCTCGCGACACCGGCACCG 27191  
RESULT 17  
AAH79278/c  
ID AAH79278 standard; DNA; 31422 BP.  
XX  
XX AAH79278;  
XX  
XX 04-DEC-2001 (first entry)  
XX  
XX Streptomyces avermitilis coding sequences SEQ ID NO: 2.  
DE  
XX  
XX Avermectin aglycone synthase; AAS; avermectin derivative;  
KW drug production; veterinary drug; pesticide; ds.  
XX  
XX Streptomyces avermitilis.  
OS  
XX

FH Key Location/Qualifiers  
CDS 1. .14646  
FT /\*tag= a  
FT /product= "AAG65266"  
FT 14824. .31422  
FT /\*tag= b  
FT /product= "AAG65267"  
XX WO200162939-A1.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-JP001381.  
XX  
XX 24-FEB-2000; 2000JP-00047405.  
XX  
XX (YOW ) KYOWA HAKKO KOGYO KK.  
PA (KITA ) KITASATO INST.  
XX  
XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
XX  
XX WPI; 2001-582053/65.  
DR P-PSDB; AAG65266, AAG65267.  
XX  
XX New modified avermectin aglycone synthase derived from Streptomyces  
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in  
PT drugs and pesticides.  
XX  
XX Disclosure; Page 103-149; 257pp; Japanese.  
XX  
XX The present invention relates to the production of modified derivatives  
CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
CC avermitilis. The activity of an acyl carrier protein (ACP), beta-ketoacyl  
CC carrier protein synthase (KS), acyltransferase (AT), beta-ketoacyl  
CC carrier protein reductase (KR), dehydratase (DH), enoyl reductase (ER)  
CC and/or thioesterase (TE) domain may be reduced or suppressed. The process  
CC can be used in the production of drugs, veterinary drugs and pesticides.  
CC The present sequence is a fragment of the S. avermitilis genome  
XX  
XX Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 U; 0 Other;  
SQ Query Match 4.0%; Score 20; DB 4; Length 31422;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 366 CCTCGCGACACCGGCACCG 385  
Db 27210 CCTCGCGACACCGGCACCG 27191  
RESULT 18  
AAT80413  
ID AAT80413 standard; DNA; 43280 BP.  
XX  
XX AAT80413;  
XX  
XX 27-FEB-1998 (first entry)  
XX  
XX Tylactone synthase gene cluster.  
XX  
XX Tylactone synthase gene cluster; tylG gene; multifunctional protein;  
KW polyketide; tylactone synthesis; antibiotic; tylosin; ss.  
XX  
XX Streptomyces fradiae.  
XX  
XX Key Location/Qualifiers  
FH 816. .14243  
FT /\*tag= a  
FT /transl\_except= (pos: 816. .818, aa: Met)  
FT /note= "ORF1 encodes protein shown in AAW22601"  
FT 14351. .19945  
FT /\*tag= b  
FT /transl\_except= (pos: 14351. .14353, aa: Met)  
FT

```
FT FT /note= "ORF2 encodes protein shown in AAW22602"  
FT CDS 20010.31199  
FT FT /*tag= c  
FT FT /transl_except= (pos: 20010..20012, aa: Met)  
FT FT /note= "ORF3 encodes protein shown in AAW22603"  
FT CDS 31232.36067  
FT FT /*tag= d  
FT FT /note= "ORF4 encodes protein shown in AAW22604"  
FT FT 36249..41774  
FT FT /*tag= e  
FT FT /note= "ORF5 encodes protein shown in AAW22605"  
FT FT  
FT XX EP791655-A2.  
FT XX  
FT XX  
FT XX 27-AUG-1997. 97EP-00301056.  
FT XX  
FT XX 19-FEB-1997; 97EP-00301056.  
FT XX  
FT XX 22-FEB-1996; 96US-0012078P.  
FT XX (ELIL ) LILLY & CO ELI.  
FT XX  
FT XX Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;  
FT XX WPT; 1997-418046/39.  
FT DR P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.  
FT XX  
FT XX DNA encoding Streptomyces fradiae tyllactone synthase domain - for  
FT FT production of tylosin-related polyketide compounds.  
FT XX  
FT PS Claim 2; Page 8-66; 220pp; English.  
FT XX  
FT XX This sequence represents the tyllactone synthase gene cluster of the  
FT CC invention. This sequence is also referred to as the tylG gene, and was  
FT CC isolated from Streptomyces fradiae. This sequence encodes multifunctional  
FT CC proteins which direct the synthesis of the polyketide tyllactone, isolated  
FT CC from Streptomyces fradiae. Tyllactone is the basic building block of the  
FT CC antibiotic tylosin. The DNA sequence can be modified so as to alter the  
FT CC type of carboxylic acids incorporated, the number of carboxylic acids  
FT CC incorporated and/or the post-condensation reactions performed, thereby  
FT CC resulting in novel tylosin-related polyketides  
FT XX  
FT SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 U; 0 Other;  
FT  
FT Query Match 4.0%; Score 20; DB 2; Length 43280;  
FT Best Local Similarity 100.0%; Pred. No. 14;  
FT Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
FT  
FT Qy 169 GCCACCGGGCTGCGGCTGCC 188  
FT Db 13893 GCCACCGGGCTGCGGCTGCC 13912  
FT  
FT RESULT 19  
FT AAL17049/c  
FT ID AAL17049 standard; cDNA; 177 BP.  
FT XX  
FT AC AAL17049;  
FT XX  
FT XX 07-DEC-2001 (first entry)  
FT DT Human breast cancer expressed polynucleotide 9506.  
FT DE Human breast cancer expressed polynucleotide 9506.  
FT XX  
FT KW Human; breast cancer; cell marker; cytostatic; ss.  
FT XX  
FT OS Homo sapiens.  
FT XX  
FT PN WO200151628-A2.  
FT XX  
FT PD 19-JUL-2001.  
FT XX  
FT PF 10-JAN-2001; 2001WO-US000798.  
FT XX  
FT XX 14-JAN-2000; 2000US-0176077P.  
FT PR 14-MAR-2000; 2000US-0189167P.  
FT PR 24-MAR-2000; 2000US-0192099P.  
FT PR 29-MAR-2000; 2000US-0193480P.  
FT PR 15-MAY-2000; 2000US-0205230P.  
FT PR 09-JUN-2000; 2000US-0211315P.  
FT PR 25-JUL-2000; 2000US-0220534P.  
FT XX  
FT PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
FT XX  
FT PI Lillie J, Xu Y, Wang Y, Steinmann K;  
FT XX WPT; 2001-451856/48.  
FT XX
```

```
PR 14-JAN-2000; 2000US-0176077P.  
PR 14-MAR-2000; 2000US-0189167P.  
PR 24-MAR-2000; 2000US-0192099P.  
PR 29-MAR-2000; 2000US-0193480P.  
PR 15-MAY-2000; 2000US-0205230P.  
PR 09-JUN-2000; 2000US-0211315P.  
PR 25-JUL-2000; 2000US-0220534P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Lillie J, Xu Y, Wang Y, Steinmann K;  
XX WPT; 2001-451856/48.  
XX  
PT New peptide useful as a marker for the diagnosis of breast cancer.  
XX  
PS Claim 1; Page 1699; 3695pp; English.  
XX  
CC The invention relates to human breast cancer expressed polynucleotides  
CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterising treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity  
XX  
SQ Sequence 177 BP; 35 A; 56 C; 49 G; 37 T; 0 U; 0 Other;  
FT  
FT Query Match 3.8%; Score 19; DB 4; Length 177;  
FT Best Local Similarity 100.0%; Pred. No. 79;  
FT Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
FT  
FT Qy 12 CGCCGAGCAGAACCGCATC 30  
FT Db 157 CGCCGAGCAGAACCGCATC 139  
FT  
FT RESULT 20  
FT AAL08770/c  
FT ID AAL08770 standard; cDNA; 177 BP.  
FT XX  
FT AC AAL08770;  
FT XX  
FT XX 07-DEC-2001 (first entry)  
FT DT Human breast cancer expressed polynucleotide 1227.  
FT DE Human breast cancer expressed polynucleotide 1227.  
FT KW Human; breast cancer; cell marker; cytostatic; ss.  
FT XX  
FT OS Homo sapiens.  
FT XX  
FT PN WO200151628-A2.  
FT XX  
FT PD 19-JUL-2001.  
FT XX  
FT PF 10-JAN-2001; 2001WO-US000798.  
FT XX  
FT XX 14-JAN-2000; 2000US-0176077P.  
FT PR 14-MAR-2000; 2000US-0189167P.  
FT PR 24-MAR-2000; 2000US-0192099P.  
FT PR 29-MAR-2000; 2000US-0193480P.  
FT PR 15-MAY-2000; 2000US-0205230P.  
FT PR 09-JUN-2000; 2000US-0211315P.  
FT PR 25-JUL-2000; 2000US-0220534P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Lillie J, Xu Y, Wang Y, Steinmann K;  
XX WPT; 2001-451856/48.  
XX
```

PT New peptide useful as a marker for the diagnosis of breast cancer.

PS Claim 1; Page 272; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides  
 CC (AA07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity

SQ Sequence 177 BP; 35 A; 56 C; 49 G; 37 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 4; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCGGAGCAGACCGCATC 30  
 |||||  
 Db 157 CGCGGAGCAGACCGCATC 139

RESULT 21

AAAT73432  
 ID AAT73432 standard; DNA; 375 BP.

XX AC AAT73432;

XX 07-OCT-1997 (first entry)

DE Human precursor urocortin coding sequence.

XX Human; precursor; urocortin; corticotropin releasing factor; receptor;  
 KW adrenocorticotrophic hormone; ACTH; affinity; rat; ovine; carp; urotensin;  
 KW p.acenta; probe; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH mat\_peptide 247..366  
 FT /\*tag= a  
 FT /\*tag= a

XX WO9700063-A2.

XX 03-JAN-1997.

XX 12-JUN-1996; 96WO-US010240.

XX 13-JUN-1995; 95US-0028444P.

XX 13-JUN-1995; 95US-00490314.

XX 11-AUG-1995; 95US-0002223P.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Vale WW, Vaughan J, Donaldson CJ, Lewis KA, Sawchenko P;

XX Rivier JEF, Perrin MH;

XX WPI; 1997-077344/07.

XX P-PSDB; AAW22392.

XX Urocortin peptide(s) related to urotensin and corticotropin-releasing

XX factor - for increasing ACTH and beta-endorphin levels, lowering blood

XX pressure and improving mood, memory and learning performance.

XX Claim 13; Page 82; 91pp; English.

XX This nucleotide sequence encodes a human precursor urocortin peptide

XX which: (i) binds to corticotropin-releasing factor (CRF) receptor 2 (CRF-

XX R2); (ii) promotes the production of adrenocorticotrophic hormone (ACTH);

CC rat/human CRF (r/hCRF); and (iv) has < 80% homology with r/hCRF, ovine  
 CC CRF or carp urotensin. The sequence was isolated from a human genomic  
 CC placental DNA library using a 160 bp fragment of the corresponding rat  
 CC gene (AAT73433) as a probe. The nucleotide sequence encoding the mature  
 CC human urocortin peptide has 88% similarity to the rat sequence (AAT73433)

XX Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 375;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCGGAGCAGACCGCATC 30  
 |||||

Db 333 CGCGGAGCAGACCGCATC 351

RESULT 22

ACA63335  
 ID ACA63335 standard; DNA; 375 BP.

XX AC ACA63335;

XX 29-AUG-2003 (first entry)

XX DNA encoding human urocortin.

XX Human; urocortin; adrenocorticotrophic hormone; ACTH; blood flow; memory;  
 KW central nervous system disorder; blood pressure; vascular bed; anxiety;  
 KW coronary blood flow; inflammation; vascular permeability; parturition;  
 KW CRF-binding protein; hypothalamic pituitary adrenal function; swelling;  
 KW Alzheimer's disease; chronic fatigue syndrome; appetite; alertness; ds;  
 KW respiratory system; learning performance; depression; endocrine disorder;  
 KW gene.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 1..375  
 FT /\*tag= a  
 FT /\*tag= a

XX US2003032587-A1.

XX 13-FEB-2003.

XX 26-MAR-2001; 2001US-00818009.

XX 13-JUN-1995; 95US-0028444P.

XX 11-AUG-1995; 95US-0002223P.

XX 12-JUN-1996; 96WO-US010240.

XX 10-DEC-1997; 97US-00981189.

XX (SALK ) SALK INST BIOLOGICAL STUDIES.

XX Valle WW, Vaughan J, Donaldson CJ, Lewis KA, Sawchenko P;

XX Rivier JEF, Perrin MH;

XX WPI; 1997-077344/07.

XX P-PSDB; ABU62366.

XX Urocortin peptide(s) related to urotensin and corticotropin-releasing

XX factor - for increasing ACTH and beta-endorphin levels, lowering blood

XX pressure and improving mood, memory and learning performance.

XX Disclosure; Page 7; 34pp; English.

XX The invention relates to a human urocortin (Ucn) peptide or an analogous

XX sequence having only conservative substitutions to the amino acid

XX residues in it, or an N-terminally shortened fragment of either which is

XX biologically active to increase adrenocorticotrophic hormone (ACTH)

CC production. Human urocortin or its N-terminally shortened antagonist

XX peptide are useful for modifying blood flow and/or blood pressure and is

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC further useful for modulating blood flow in a desired vascular bed. Human  
 CC urocortin is also useful for increasing coronary blood flow and for  
 CC decreasing swelling and/or inflammation and/or vascular permeability. A  
 CC CRF-binding protein blocking compound is useful for increasing the in  
 CC vivo level of CRF and/or Ucn. The amount of CRF-binding protein blocking  
 CC compound is sufficient to promote parturition in a pregnant female. The  
 CC amount of the compound administered is effective so as to result in an  
 CC increase in free endogenous CRF and/or Ucn in the brain which causes  
 CC improvement in short to medium term memory in a subject afflicted with  
 CC Alzheimer's disease, relief from chronic fatigue syndrome, suppression of  
 CC appetite, stimulation of the respiratory system, improvement in learning  
 CC performance, improvement in memory, improvement in alertness, reduction  
 CC of depression and/or lessening of anxiety. The compound is administered  
 CC so that it reaches the brain. Human urocortin is useful for evaluating  
 CC hypothalamic pituitary adrenal function in mammals with suspected  
 CC endocrine or central nervous system pathology. Human urocortin antibodies  
 CC are useful in diagnostic methods and systems for detecting the level of  
 CC Ucn polypeptide, for immunoaffinity or affinity chromatography  
 CC purification of Ucn, and also in human therapeutic methods. The present  
 CC sequence represents DNA encoding human urocortin

SQ Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATC 30  
 |||||  
 Db 333 CGCCGAGCAGACCGCATC 351

## RESULT 23

AAZ35707  
 ID AAZ35707 standard; cDNA; 375 BP.

AC AAZ35707;

DT 31-JAN-2000 (first entry)

DE Human urocortin encoding cDNA.

XX Human; urocortin; protection; ischaemic injury; cardiac myocyte;  
 KW cardioprotective; Cardiac ischaemia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH 1..375  
 CDS /\*tag= a  
 FT /product= "urocortin"

FN WO951261-A1.

PD 14-OCT-1999.

XX 07-APR-1999; 99WO-GH001056.

XX 07-APR-1998; 98GB-00007504.

XX (UNLO ) UNIV COLLEGE LONDON.

XX Stephanou A, Brar B, Knight R, Latchman D;

XX WPI; 2000-013030/01.

XX P-PSDB; AAY49943.

XX Use of urocortin or cardio-protective derivative, useful for preparation  
 PT of medicament for treating cardiac ischemia.

XX Disclosure; Page 35-36; 42pp; English.

XX The present sequence encodes human urocortin. The present invention

CC describes the use of urocortin or its cardio-protective derivative, in  
 CC the manufacture of a medicament for use in the treatment of cardiac  
 CC ischaemia. Also described are: (1) a method for the treatment of a  
 CC patient who has suffered a cardiac ischaemia, comprising administering  
 CC urocortin to the patient; (2) urocortin for use in a method of treatment  
 CC of the human or animal body by therapy; and (3) an agent for treating  
 CC cardiac ischaemia comprising urocortin

XX Sequence 375 BP; 52 A; 128 C; 145 G; 50 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 3; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATC 30  
 |||||  
 Db 333 CGCCGAGCAGACCGCATC 351

## RESULT 24

AAV99120  
 ID AAV99120 standard; DNA; 452 BP.

XX AAV99120;

XX 09-MAR-1999 (first entry)

XX DNA methyltransferase genomic fragment.

XX DNA methyltransferase; DNA Mefase; antisense oligonucleotide; human;  
 KW cellular growth; tumour growth inhibition; silenced gene activation;  
 KW beta thalassemia; sickle cell anemia; ss.

XX Homo sapiens.

XX WO9854313-A2.

XX 03-DEC-1998.

XX 29-MAY-1998; 98WO-IB001107.

XX 30-MAY-1997; 97US-00866340.

XX 17-DEC-1997; 97US-0069865P.

XX (UYMC-) UNIV MCGILL.

XX Szyf M, Bigey P, Ramchandani S;

XX WPI; 1999-059833/05.

XX New DNA methyltransferase nucleotide sequences - used particularly to  
 PT develop antisense oligonucleotides for diagnostic and therapeutic  
 PT purposes, particularly for inhibiting tumour growth.

XX Claim 1; Page 59; 108pp; English.

XX AAV99091-128 represent DNA methyltransferase (DNA Mefase) genomic  
 CC fragments. The sequences together sequentially produce the full length  
 CC sequence AAV99128. The sequences can be used to derive antisense  
 CC oligonucleotides which inhibit DNA Mefase expression. The antisense  
 CC oligonucleotides can be used in investigating the role of DNA Mefase in  
 CC cellular growth. They can be administered at different points in the cell  
 CC cycle, or in conjugation with promoters or inhibitors of cell growth to  
 CC determine the role of DNA Mefase in the growth of the cell type of  
 CC interest. The antisense oligonucleotides can also be used for inhibiting  
 CC tumour growth in a mammal, or to activate silenced genes to provide a  
 CC missing gene function. This ameliorates disease symptoms, e.g. in beta  
 CC thalassemia and sickle cell anemia. The antisense oligonucleotides can  
 CC also be used as analytical and diagnostic tools and a potentiators of  
 CC transgenic plant and animal studies

XX Sequence 452 BP; 100 A; 135 C; 131 G; 86 T; 0 U; 0 Other;

```
Query Match      3.8%; Score 19; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels
```

Qy 27 CATCCTGCTGAAGCTGGTC 45  
|||  
Db 201 CATCCTGCTGAAGCTGGTC 219

RESULT 25  
ABS63429  
ID ABS63429 standard; cDNA; 953 BP.  
XX  
XX ABS63429; AC  
XX  
XX  
DT 15-NOV-2002 (first entry)  
XX  
XX  
DE DNA encoding wheat coffeoyl-CoA O-methyltransferase (CCOMT).  
XX  
XX Plant; coffeoyl-CoA O-methyltransferase; CCOMT; transgenic; corn; rice;  
KW liginin; sovbear; wheat; gene; ss

Novel plant caffeoyl-CoA O-methyltransferase polypeptide and polynucleotide encoding it useful for producing transgenic plants having altered level of lignin production in cells.

xx PS Claim 2: Page 48-49; 59pp; English.

The invention relates to novel isolated plant caffeoyl-CoA O-methyltransferase (CCOMT) polypeptides (I) and encoding polynucleotides (II). (I) is useful for evaluating at least one compound for its ability to inhibit the activity of a caffeoyl-CoA O-methyltransferase. (II) is useful as a probe or primer for obtaining a nucleic acid fragment encoding CCOMT polypeptide by hybridisation and amplification. (II) or a polynucleotide comprising at least 30 contiguous nucleotides derived from (II), is useful for selecting an isolated polynucleotide that affects the level of expression of methyltransferase polypeptide in a plant cell. (II) is useful for positive selection of a transformed cell, by transforming a plant cell, preferably a monocot or dicot, with (II) and growing the transformed plant cell under conditions allowing expression of the polynucleotide to complement CCOMT auxotroph to provide a positive selection. (I) is useful for preparing antibodies for use in detecting CCOMT polypeptides *in situ* in cells or *in vitro* in cell extracts. (II) or its fragments are useful as probes for genetically and physically mapping the genes that they are part of and as markers for traits linked to those genes. Fragments of (II) are useful to create transgenic plants in which CCOMT polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This is useful for altering the level of lignin production in the plant cells. (II) is also useful in direct fluorescence *in situ* hybridisation mapping. ABS63407-ABS63430 represent plant CCOMT coding sequences of the invention.

Sequence 953 BP: 192 A; 336 C; 279 G; 146 T; 0 U; 0 Other;  
XX SO

Query Match	3.8%	Score 19;	DB 6;	Length 953;
Best Local Similarity	100.0%;	Fred. No. 64;		
Matches	19;	Conservative	0; Mismatches	0; Indels
Gaps	0;			

RESULT 26  
ABS63428  
ID ABS63428 standard; cDNA; 1049 BP.  
XX  
XX ABS63428;  
XX  
XX  
DT 15-NOV-2002 (first entry)  
XX  
XX DNA encoding wheat caffeoyl-CoA O-methyltransferase (CCONT).  
DE  
DE Plant; caffeoyl-CoA O-methyltransferase; CCONT; transgenic; corn; rice;  
KW lignin; soybean; wheat; gene; ss.  
KW

XX Novel plant caffeoyl-CoA O-methyltransferase polypeptide and  
PT polynucleotide encoding it useful for producing transgenic plants having  
PT altered level of lignin production in cells.

XX  
PS Claim 2: Page 47: 59pp: English.

The invention relates to novel isolated plant caffeoyl-CoA O-methyltransferase (CCOMT) polypeptides (I) and encoding polynucleotides (II). (I) is useful for evaluating at least one compound for its ability to inhibit the activity of a caffeoyl-CoA O-methyltransferase. (II) is useful as a probe or primer for obtaining a nucleic acid fragment encoding CCOMT polypeptide by hybridisation and amplification. (II) or a polynucleotide comprising at least 30 contiguous nucleotides derived from (II), is useful for selecting an isolated polynucleotide that affects the level of expression of methyltransferase polypeptide in a plant cell. (II) is useful for positive selection of a transformed cell, by transforming a plant cell, preferably a monocot or dicot, with (II) and growing the transformed plant cell under conditions allowing expression of the polynucleotide to complement CCOMT auxotroph to provide a positive selection. (I) is useful for preparing antibodies for use in detecting CCOMT polypeptides *in situ* in cells or *in vitro* in cell extracts. (II) or its fragments are useful as probes for genetically and physically mapping the genes that they are part of and as markers for traits linked to those genes. Fragments of (II) are useful to create transgenic plants in which CCOMT polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This is useful for altering the level of lignin production in the plant cells. (II) is also useful in direct fluorescence *in situ* hybridisation mapping. ABS63407-ABS63430 represent plant CCOMT coding sequences of the invention.

30X



```
SQ Sequence 1049 BP; 207 A; 341 C; 303 G; 191 T; 0 U; 7 Other;
Query Match 3.8%; Score 19; DB 6; Length 1049;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGTCCTCGG 49
Db 591 CTGCTGAAGCTGTCCTCGG 609

RESULT 27
AAQ82825
ID AAQ82825 standard; cDNA; 1368 BP.
XX
XX
AC AAQ82825;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-SEP-1995 (first entry)
XX
XX cDNA encoding an enzyme with endoglucanase activity.
DE endoglucanase; degradation; modification; plant material; ss.
XX
XX Hypocrea lixii.
XX
XX WO9502043-A1.
XX
PD 19-JAN-1995.
XX
PF 05-JUL-1994; 94WO-DK000275.
XX
XX 06-JUL-1993; 93DK-00000812.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Dalboge H, Christgau S, Andersen LN, Kofod LV, Kauppinen MS;
XX
XX WPI; 1995-066891/09.
XX
XX Enzyme from Trichoderma harzianum having endoglucanase activity - used in
PT degrading or modifying plant cell walls e.g. in food or feed prepn.
XX
XX Claim 1; Page 25-26; 33pp; English.
XX
XX This cDNA encodes a novel protein with endoglucanase activity isolated
CC from Trichoderma harzianum. Endoglucanases constitute a gp. of
CC hydrolases, which catalyse endo hydrolysis of 1,4-beta-D- glycosidic
CC linkages in cellulose and its derivs., lichenin, beta-1,4 bonds in mixed
CC beta-1,3 glucans and other plant material contg. cellulosic parts. The
CC enzyme can be used for the degradation of plant cell wall tissue.
CC (Updated on 25-MAR-2003 to correct FN field.) (Updated on 16-OCT-2003 to
CC standardise OS field)
XX
XX Sequence 1368 BP; 327 A; 412 C; 307 G; 322 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1368;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 CGTGAACGAGCCCTCGG 489
Db 813 CGTGAACGAGCCCTCGG 831

RESULT 28
AAX86613/C
ID AAX86613 standard; cDNA; 1698 BP.
XX
XX AAX86613;
XX
XX 15-OCT-1999 (first entry)

cDNA encoding an acetylcholinesterase protein.
Acetylcholinesterase; AchE; fusion protein; ligand receptor; monomer;
ligand detection; marker enzyme; ss.
Bungarus fasciatus.

Key Location/Qualifiers
CDS 1..1698
FT /*tag=a
FT /transl_except= (pos: 52..54, aa: Leu)
FT /transl_except= (pos: 55..57, aa: Cys)
FT /transl_except= (pos: 58..60, aa: Ile)
FT /transl_except= (pos: 61..63, aa: Pro)
FT /transl_except= (pos: 64..66, aa: Ser)
FT /transl_except= (pos: 67..69, aa: Cys)
FT /transl_except= (pos: 70..72, aa: Val)
FT /transl_except= (pos: 73..75, aa: Ala)
FT /transl_except= (pos: 76..78, aa: Val)
FT /transl_except= (pos: 79..81, aa: Leu)
FT /transl_except= (pos: 82..84, aa: Gln)
FT /transl_except= (pos: 85..87, aa: Leu)
FT /product= "acetylcholinesterase"
FT /note= "no termination codon given"
XX
XX FR2773802-A1.
XX
XX 23-JUL-1999.
XX
XX 22-JAN-1998; 98FR-00000656.
XX
XX 22-JAN-1998; 98FR-00000656.
XX
XX (INSP) INST PASTEUR.
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bon C, Choumet V, Cousin X;
XX
XX WPI; 1999-471239/40.
XX P-PSDB; AAY30100.
XX
XX A fusion protein comprising an acetyl cholinesterase and ligand receptor,
PT useful for detection of ligands.
XX
XX Claim 19; Page 98-99; 114pp; French.
XX
XX The present sequence encodes an acetylcholinesterase protein (AChE). The
CC protein is used to make a fusion protein comprising an AChE monomer and a
CC specific ligand receptor. The AChE fusion protein is useful for the
CC production of an AChE monomer in a soluble format. The AChE fusion
CC polypeptide is useful for detection of ligands in samples. AChE is used
CC as a marker enzyme, in a similar manner to peroxidase, alkaline
CC phosphatase and beta-galactosidase. By having AChE fused to a receptor
CC protein, various ligands can be detected by their binding to the receptor
CC portion of the fusion polypeptide
XX
XX Sequence 1698 BP; 305 A; 595 C; 495 G; 303 T; 0 U; 0 Other;

Query Match 3.8%; Score 19; DB 2; Length 1698;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 CCACCGAGCGGCCCGGC 425
Db 712 CCACCGAGCGGCCCGGC 694

RESULT 29
AAS54382
ID AAS54382 standard; DNA; 1893 BP.
XX
XX AAS54382;
XX
```

DT 13-FEB-2002 (first entry)  
XX Pseudomonas aeruginosa DNA for cellular proliferation protein #513.  
DE Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
XX antibacterial; drug design.  
KW Pseudomonas aeruginosa.  
XX WO200170955-A2.  
XX 27-SEP-2001.  
XX 21-MAR-2001; 2001WO-US003180.  
XX 21-MAR-2000; 2000US-0191078P.  
XX 23-MAY-2000; 2000US-0206848P.  
XX 26-MAY-2000; 2000US-0207727P.  
XX 23-OCT-2000; 2000US-0242578P.  
XX 27-NOV-2000; 2000US-0253625P.  
XX 22-DEC-2000; 2000US-0257931P.  
XX 16-FEB-2001; 2001US-0269308P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX P-PSDB; AAU36523.  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
XX Claim 27; SEQ ID NO 8019; 511pp; English.  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence encodes an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1893 BP; 374 A; 627 C; 575 G; 317 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 4; Length 1893;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 84 CGAAGGCATCGCCGCGC 102  
Db 807 CGAAGGCATCGCCGCGC 925  
RESULT 30  
ACA42802  
ID ACA42802 standard; DNA; 1893 BP.  
XX AC ACA42802;  
XX

DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #24459.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Pseudomonas aeruginosa.  
XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU38932.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 30672; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1893 BP; 374 A; 627 C; 575 G; 317 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 7; Length 1893;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CGAAGGATCGCCCGCGC 102  
 DB 807 CGAAGGATCGCCCGCGC 825  
 RESULT 31  
 ABN85313/c  
 ID ABN85313 standard; cDNA; 2040 BP.  
 XX AC ABN85313;  
 XX DT 30-SEP-2002 (first entry)  
 XX DE Human cytoskeleton-associated protein, CSAP-4, coding sequence.  
 XX KW Human; cytoskeleton-associated protein; CSAP; CSAP-4;  
 KW cell proliferative disorder; viral infection; neurological disorder;  
 KW transgenic animal; antiatherosclerotic; antiproliferative; antiinflammatory;  
 KW virucide; anticonvulsant; vasotropic; cerebroprotective; neurotropic;  
 KW neuroprotective; cytostatic; gene therapy; gene; ss.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 XX CDS 291..1271  
 XX FT /\*tag= a  
 XX ET /product= "CSAP-4"  
 XX FN WO200253719-A2.  
 XX PD 11-JUL-2002.  
 XX PF 04-JAN-2002; 2002WO-US000178.  
 XX PR 04-JAN-2001; 2001US-0260085P.  
 XX PR 13-FEB-2001; 2001US-0268554P.  
 XX PR 14-FEB-2001; 2001US-0269111P.  
 XX PR 23-FEB-2001; 2001US-0271211P.  
 XX PA (INCY-) INCYTE GENOMICS INC.  
 XX PI Lu DAM, Baughn MR, Yao MG, Ding L, Honchell CD, Yue H, Tang YT;  
 PI Warren BA, Duggan BM, Xu Y, Wallia NK, Griffin JA, Stewart BA;  
 PI Gandhi AR, Khan FA, Thangavelu K, Ison CH, Azimzai Y, Hafalia AJA;  
 PI Gietzen KJ, Lal PG, Sanjanwala NM, Elliott VS;  
 XX WPI; 2002-583611/62.  
 XX DR P-PSDB; ABB83475.  
 XX PT Novel isolated human cytoskeleton-associated protein for diagnosing,  
 PT treating or preventing atherosclerosis, psoriasis, leukemia, epilepsy,  
 PT ischemic cerebrovascular disease, cerebral neoplasms and Alzheimer's  
 PT disease.  
 XX PS Claim 5; Page 154-155; 167pp; English.  
 XX CC The present sequence is the coding sequence for a human cytoskeleton-  
 CC associated protein (CSAP). The CSAP and its coding sequence are useful in  
 CC the diagnosis, treatment and prevention of a cell proliferative disorder  
 CC such as actinic keratosis, atherosclerosis, psoriasis, primary  
 CC thrombocythaemia, leukaemia; a viral infection such as those caused by  
 CC adenoviruses (acute respiratory disease, pneumonia), arenaviruses  
 CC (lymphocytic choriomeningitis); and a neurological disorder such as  
 CC epilepsy, ischaemic cerebrovascular disease, stroke, cerebral neoplasms,  
 CC Alzheimer's disease, Pick's disease, Huntington's disease or amyotrophic  
 CC lateral sclerosis. The CSAP coding sequence is also useful for creating  
 CC knock out or knock in humanised animals or transgenic animals to model  
 CC human diseases  
 XX SQ Sequence 2040 BP; 465 A; 584 C; 650 G; 341 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 6; Length 2040;  
 Best Local Similarity 100.0%; Pred. No. 59;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 12 CGCCGAGCAGAACCGCATC 30  
 DB 1949 CGCCGAGCAGAACCGCATC 1931  
 RESULT 32  
 ADD29815  
 ID ADD29815 standard; mRNA; 2448 BP.  
 XX AC ADD29815;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human tumour suppressor mRNA SEQ ID NO:311.  
 XX KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.  
 XX OS Homo sapiens.  
 XX FN WO20003058201-A2.  
 XX PD 17-JUL-2003.  
 XX PF 31-DEC-2002; 2002WO-US041825.  
 XX PR 31-DEC-2001; 2001US-0345317P.  
 XX PA (QUAR-) QUARK BIOTECH INC.  
 XX PA (CLEV-) CLEVELAND CLINIC FOUND.  
 XX PI Feinstein E, Gudkov AV;  
 XX WPI; 2003-598393/56.  
 XX PT Diagnosing cancer comprises determining the polypeptide or polynucleotide  
 PT levels e.g., hepatic lipase, in a sample from a subject, where a higher  
 PT level compared to that in a subject free of cancer is indicative of  
 PT cancer.  
 XX PS Disclosure; SEQ ID NO 311; 272pp; English.  
 XX CC The invention relates to a novel method for diagnosing a cancer in a  
 CC subject. The method comprises determining, in a sample from the subject,  
 CC the level of at least one polypeptide, where a higher level of the  
 CC polypeptide compared to the level of the polypeptide in a subject free of  
 CC cancer is indicative of cancer. The polypeptide is selected from any of  
 CC the polypeptides encoded by the polynucleotides listed in the  
 CC specification and polypeptides which are at least 70% homologous to the  
 CC polypeptides. The method of the invention has cytostatic activity, and  
 CC may have a use in gene therapy. The method is useful in identifying  
 CC markers specific for one or several types of cancer, depending on the  
 CC tissue origin, which may be used in numerous diagnostic and prognostic  
 CC applications as well as cancer type-specific targets for therapeutic  
 CC intervention. The compounds that modulate the activity of a tumour  
 CC suppressor gene are useful in the treatment of cancer or as anti-cancer  
 CC drugs. The present sequence represents a polynucleotide of the invention.  
 XX SQ Sequence 2448 BP; 573 A; 728 C; 609 G; 522 T; 0 U; 16 Other;  
 Query Match 3.8%; Score 19; DB 9; Length 2448;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 418 GCCCGCGCGCGCGAGCAGC 436  
 DB 315 GCCCGCGCGCGCGAGCAGC 333  
 RESULT 33  
 AAX86614/c  
 ID AAX86614 standard; cDNA; 2478 BP.

```

XX AC AAX86614;
XX DT 15-OCT-1999 (first entry)
XX DE cDNA encoding an acetylcholinesterase protein/scFv fusion protein.
XX KW Acetylcholinesterase; AChE; fusion protein; ligand receptor; monomer;
XX KW ligand detection; marker enzyme; ss.
XX OS Synthetic.
XX OS Bungarus fasciatus.
XX FH Key
XX FT CDS
XX FT 1. .2478
XX FT /tag= a
XX FT /transl_except= (pos: 52. .54, aa: Leu)
XX FT /transl_except= (pos: 55. .57, aa: Cys)
XX FT /transl_except= (pos: 58. .60, aa: Ile)
XX FT /transl_except= (pos: 61. .63, aa: Pro)
XX FT /transl_except= (pos: 64. .66, aa: Ser)
XX FT /transl_except= (pos: 67. .69, aa: Cys)
XX FT /transl_except= (pos: 70. .72, aa: Val)
XX FT /transl_except= (pos: 73. .75, aa: Ala)
XX FT /transl_except= (pos: 76. .78, aa: Val)
XX FT /transl_except= (pos: 79. .81, aa: Leu)
XX FT /transl_except= (pos: 82. .84, aa: Gln)
XX FT /transl_except= (pos: 85. .87, aa: Leu)
XX FT /note= "no termination codon given"
XX PN FR2773802-A1.
XX XX
XX PD 23-JUL-1999.
XX PF 22-JAN-1998; 98FR-00000656.
XX PR 22-JAN-1998; 98FR-00000656.
XX XX
XX PA (INSP ) INST PASTEUR.
XX PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX XX
XX PI Bon C, Choumet V, Cousin X;
XX DR WPI: 1999-471239/40.
XX DR P-PSDB; AAY30101.
XX XX
XX FT A fusion protein comprising an acetyl cholinesterase and ligand receptor,
XX FT useful for detection of ligands.
XX PS Claim 20; Page 99-102; 114pp; French.
XX CC The present sequence encodes an acetylcholinesterase protein (AChE)/ scFv
XX CC fusion protein of the invention. The specification describes a fusion
XX CC protein comprising an AChE monomer and a specific ligand receptor. The
XX CC AChE fusion protein is useful for the production of an AChE monomer in a
XX CC soluble format. The AChE fusion polypeptide is useful for detection of
XX CC ligands in samples. AChE is used as a marker enzyme, in a similar manner
XX CC to peroxidase, alkaline phosphatase and beta-galactosidase. By having
XX CC AChE fused to a receptor protein, various ligands can be detected by
XX CC their binding to the receptor portion of the fusion polypeptide
XX XX
SQ Sequence 2478 BP; 494 A; 796 C; 710 G; 478 T; 0 U; 0 Other;
Query Match 3.8%; Score 19; DB 2; Length 2478;
Best Local Similarity 100.0%; Pred.No. 58;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 407 CCACCGAGCGGCCCGGC 425
Db 712 CCACCGAGCGGCCCGGC 694
RESULT 34

```

```

ABQ88107
ID ABQ88107 standard; cDNA; 3081 BP.
XX AC ABQ88107;
XX DT 18-SEP-2002 (first entry)
XX DE Human osteoblast differentiation related cDNA SEQ ID NO 14.
XX KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
XX KW osteoporosis; osteopathic; ss.
XX OS Homo sapiens.
XX PN WO200250301-A2.
XX PD 27-JUN-2002.
XX PF 18-DEC-2001; 2001WO-US048276.
XX PR 18-DEC-2000; 2000US-0255882P.
XX PR 24-APR-2001; 2001US-0285691P.
XX XX
XX PA (GENE-) GENE LOGIC INC.
XX PA (PROC ) PROCTER & GAMBLE CO.
XX XX
XX PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
XX PI Mertz L;
XX XX
XX DR WPI: 2002-557663/59.
XX XX
XX PT Use of genes and their expression profiles associated with osteoblast
XX PT differentiation for screening modulators bone formation, for diagnosing
XX PT or treating e.g. osteoporosis, or as markers for the differentiation
XX PT process.
XX PS Claim 1; SEQ ID NO 14; 78pp + Sequence Listing; English.
XX XX
XX CC The invention relates to genes and their expression profiles are used
XX CC for: (a) screening modulators of precursor stem cell differentiation into
XX CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
XX CC deposition of bone tissue, abnormal rate of osteoblast formation or
XX CC osteoporosis; or (c) treating or monitoring of bone tissue deposition.
XX CC cited in (b), or monitoring the progression of bone tissue deposition.
XX CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
XX CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
XX CC induced abnormalities in bone formation or bone loss, conditions that
XX CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
XX CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
XX CC or fibrous dysplasia. The present sequence is that of an osteoblast
XX CC differentiation associated cDNA marker of the invention. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX XX
SQ Sequence 3081 BP; 613 A; 889 C; 882 G; 697 T; 0 U; 0 Other;
Query Match 3.8%; Score 19; DB 6; Length 3081;
Best Local Similarity 100.0%; Pred.No. 56;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 CGCGAGCAGACCGCATC 30
Db 2915 CGCGAGCAGACCGCATC 2933
RESULT 35
AAI59458
ID AAI59458 standard; cDNA; 5092 BP.
XX AC AAI59458;
XX DT 22-OCT-2001 (first entry)

```

XX DE Human polynucleotide SEQ ID NO 1661.  
XX DE  
XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX OS Homo sapiens.  
XX PN WO200153312-A1.  
XX PD 26-JUL-2001.  
XX PF 26-DEC-2000; 2000WO-US034263.  
XX PR 23-DEC-1999; 99US-00471275.  
XX PR 21-JAN-2000; 2000US-00488725.  
XX PR 25-APR-2000; 2000US-00552317.  
XX PR 20-JUN-2000; 2000US-00598042.  
XX PR 19-JUL-2000; 2000US-00620312.  
XX PR 03-AUG-2000; 2000US-00653450.  
XX PR 14-SEP-2000; 2000US-00662191.  
XX PR 19-OCT-2000; 2000US-00693036.  
XX PR 29-NOV-2000; 2000US-00727344.  
XX PA (HYSEQ-) HYSEQ INC.  
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou F, Goodrich R, Drmanac RI;  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM40302.  
XX DR  
XX PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX PS  
XX PS Claim 1; SEQ ID NO 1661; 10078pp; English.  
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nontropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX SQ Sequence 5092 BP; 1310 A; 1441 C; 1448 G; 893 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 4; Length 5092;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 27 CATCTGCTGAGCTGTC 45  
Db 3813 CATCTGCTGAGCTGTC 3831  
RESULT 36  
ABL62902  
ID ABL62902 standard; DNA; 5434 BP.  
XX

AC ABL62902;  
XX DT 15-MAY-2002 (first entry)  
XX DE Breast cancer related gene sequence SEQ ID NO:1239.  
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX OS Homo sapiens.  
XX PN WO200194629-A2.  
XX PD 13-DEC-2001.  
XX PF 30-MAY-2001; 2001WO-US010838.  
XX PR 05-JUN-2000; 2000US-0209473P.  
XX PR 18-SEP-2000; 2000US-0209531P.  
XX PR 18-SEP-2000; 2000US-0233133P.  
XX PR 20-SEP-2000; 2000US-0233617P.  
XX PR 20-SEP-2000; 2000US-0234009P.  
XX PR 20-SEP-2000; 2000US-0234034P.  
XX PR 22-SEP-2000; 2000US-0234052P.  
XX PR 22-SEP-2000; 2000US-0234509P.  
XX PR 22-SEP-2000; 2000US-0234567P.  
XX PR 25-SEP-2000; 2000US-0234923P.  
XX PR 25-SEP-2000; 2000US-0234924P.  
XX PR 25-SEP-2000; 2000US-0235077P.  
XX PR 25-SEP-2000; 2000US-0235082P.  
XX PR 25-SEP-2000; 2000US-0235134P.  
XX PR 25-SEP-2000; 2000US-0235280P.  
XX PR 26-SEP-2000; 2000US-0235637P.  
XX PR 27-SEP-2000; 2000US-0235711P.  
XX PR 27-SEP-2000; 2000US-0235720P.  
XX PR 27-SEP-2000; 2000US-0235840P.  
XX PR 27-SEP-2000; 2000US-0235863P.  
XX PR 28-SEP-2000; 2000US-0236028P.  
XX PR 28-SEP-2000; 2000US-0236032P.  
XX PR 28-SEP-2000; 2000US-0236033P.  
XX PR 28-SEP-2000; 2000US-0236034P.  
XX PR 28-SEP-2000; 2000US-0236109P.  
XX PR 28-SEP-2000; 2000US-0236111P.  
XX PR 29-SEP-2000; 2000US-0236842P.  
XX PR 29-SEP-2000; 2000US-0236891P.  
XX PR 02-OCT-2000; 2000US-0237172P.  
XX PR 02-OCT-2000; 2000US-0237173P.  
XX PR 02-OCT-2000; 2000US-0237278P.  
XX PR 02-OCT-2000; 2000US-0237294P.  
XX PR 02-OCT-2000; 2000US-0237295P.  
XX PR 03-OCT-2000; 2000US-0237316P.  
XX PR 03-OCT-2000; 2000US-0237425P.  
XX PR 03-OCT-2000; 2000US-0237598P.  
XX PR 03-OCT-2000; 2000US-0237604P.  
XX PR 03-OCT-2000; 2000US-0237606P.  
XX PR 03-OCT-2000; 2000US-0237608P.  
XX PR 01-NOV-2000; 2000US-0244867P.  
XX PR 01-NOV-2000; 2000US-0245084P.  
XX PA (AVAL-) AVALON PHARM.  
XX PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX

PS Claim 1; SEQ ID NO 1239; 44pp; English.  
XX The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour  
SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 6; Length 5434;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 27 CATCTGCTGAAGCTGGTC 45  
Db 3813 CATCTGCTGAAGCTGGTC 3831  
RESULT 37  
ABL63111  
ID ABL63111 standard; DNA; 5434 BP.  
XX AC ABL63111;  
XX DT 15-MAY-2002 (first entry)  
XX DE Breast cancer related gene sequence SEQ ID NO:1448.  
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
KW gene; ds.  
XX OS Homo sapiens.  
XX PN WO200194629-A2.  
XX PD 13-DEC-2001.  
XX PF 30-MAY-2001; 2001WO-US010838.  
XX PR 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 27-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 28-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
XX PA (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.  
XX Claim 1; SEQ ID NO 1448; 44pp; English.  
XX The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour  
SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 6; Length 5434;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 27 CATCTGCTGAAGCTGGTC 45  
Db 3813 CATCTGCTGAAGCTGGTC 3831  
RESULT 38  
ABL64407  
ID ABL64407 standard; DNA; 5434 BP.  
XX AC ABL64407;  
XX DT 15-MAY-2002 (first entry)

XX Stomach cancer related gene sequence SEQ ID NO:2744.  
 DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX Homo sapiens.  
 OS  
 XX WO200194629-A2.  
 PN  
 XX 13-DEC-2001.  
 PD  
 XX 30-MAY-2001; 2001WO-US010838.  
 XX  
 XX 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237315P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237538P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 XX (AVAL-) AVALON PHARM.  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 DR  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 XX agent to be tested for anti-neoplastic activity, and determining a change  
 XX in expression of a gene of a signature gene set.  
 XX  
 XX Claim 1; SEQ ID NO 2744; 44pp; English.  
 XX  
 XX The present invention describes a method (M1) for screening for an anti-

CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABU61664  
 CC to ABU70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
 CC tumour  
 XX  
 SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
 Query Match 3.8%; Score 19; DB 6; Length 5434;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 CATCTGCTGAAGCTGTC 45  
 Db 3813 CATCTGCTGAAGCTGTC 3831  
 RESULT 39  
 ID APT33842 standard; cDNA; 5434 BP.  
 XX  
 AC APT33842;  
 XX  
 DT 29-MAY-2003 (first entry)  
 XX  
 XX DNMT1 cDNA sequence SEQ ID No 1.  
 XX Cytostatic; methyl transferase inhibitor; DNA methyl transferase isoform;  
 XX gene therapy; anti-DNA methyl transferase oligonucleotide; inhibitor;  
 XX cell proliferation; neoplasia; DNMT 1; enzyme; ss.  
 XX Unidentified.  
 OS  
 XX WO200291926-A2.  
 PN  
 XX 21-NOV-2002.  
 XX  
 XX 13-MAY-2002; 2002WO-IB003120.  
 XX  
 XX 11-MAY-2001; 2001US-0290202P.  
 XX 11-MAY-2001; 2001US-0290212P.  
 XX (METH-) METHYLGENE INC.  
 XX Macleod AR;  
 XX WPI; 2003-149369/14.  
 XX  
 XX New inhibitors of DNA methyl transferase isoforms, e.g. anti-DNA methyl  
 XX transferase oligonucleotides or small molecule inhibitors of DNA methyl  
 XX transferase, useful for treating cell proliferative and differentiation  
 XX disorders.  
 XX  
 XX Claim 5; Fig 1B; 76pp; English.  
 XX  
 XX The invention relates to an agent that inhibits one or more specific DNA  
 XX methyl transferase isoforms (but not all DNA methyl transferase  
 XX isoforms), such as an anti-DNA methyl transferase oligonucleotide or a  
 XX small molecule inhibitor of DNA methyl transferase. The agents,  
 XX oligonucleotides, inhibitors and methods are useful for identifying  
 XX specific inhibition of specific DNA methyl transferase isoforms involved  
 XX in cell proliferation and/or differentiation, and thus providing a

CC treatment for cell proliferative and/or differentiation disorders, e.g.  
CC neoplasia. This polynucleotide sequence represents a DNMT1 cDNA sequence  
CC relating to the invention  
XX  
SQ Sequence 5434 BP; 1418 A; 1488 C; 1511 G; 1017 T; 0 U; 0 Other;  
Query Match 3.8%; Score 19; DB 7; Length 5434;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 27 CATCTGCTGAAGCTGGTC 45  
Db 3813 CATCTGCTGAAGCTGGTC 3831

RESULT 40  
AAS78583  
ID AAS78583 standard; cDNA; 5794 BP.  
XX  
AC AAS78583;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #14387.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG14396.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 14387; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5794 BP; 1466 A; 1580 C; 1617 G; 1128 T; 0 U; 3 Other;  
Query Match 3.8%; Score 19; DB 5; Length 5794;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 27 CATCTGCTGAAGCTGGTC 45  
Db 3829 CATCTGCTGAAGCTGGTC 3847

RESULT 41  
AAV99129  
ID AAV99129 standard; DNA; 19440 BP.  
XX  
AC AAV99129;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE DNA methyltransferase genomic sequence.  
XX  
KW DNA methyltransferase; DNA Mefase; antisense oligonucleotide; human;  
KW cellular growth; tumour growth inhibition; silenced gene activation;  
KW beta thalassemia; sickle cell anemia; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 115  
FT /tag= a  
FT /note= "this nucleotide is not present in AAV99092"  
FT 6741  
FT /tag= b  
FT /note= "this nucleotide is not present in AAV99094"  
FT 6744..6745  
FT /tag= c  
FT /note= "approximately 600 unspecified nucleotides are  
FT present between these bases"  
FT 7097..7098  
FT /tag= d  
FT /note= "approximately 1000 unspecified nucleotides are  
FT present between these bases"  
FT 7556..7557  
FT /tag= e  
FT /note= "approximately 1500 unspecified nucleotides are  
FT present between these bases"  
FT 7866..7867  
FT /tag= f  
FT /note= "approximately 700 unspecified nucleotides are  
FT present between these bases"  
FT 8542..8543  
FT /tag= g  
FT /note= "approximately 1000 unspecified nucleotides are  
FT present between these bases"  
FT 8807..8808  
FT /tag= h  
FT /note= "1700 unspecified nucleotides are present between  
FT these bases"  
FT 9099..9100  
FT /tag= i  
FT /note= "570 unspecified nucleotides are present between  
FT these bases"  
FT 9106..9108  
FT /tag= j  
FT /note= "these nucleotides are not present in AAV99103"  
FT 9355..9356  
FT /tag= k  
FT /note= "350 unspecified nucleotides are present between  
FT these bases"  
FT 9662..9663  
FT /tag= l  
FT



FT		/note= "300 unspecified nucleotides are present between these bases"
FT		11407. .11409
FT	misc_feature	/*tag= m
FT		/note= "approximately 450 unspecified nucleotides are present between these bases"
FT		11704. .11705
FT	misc_feature	/*tag= n
FT		/note= "approximately 900 unspecified nucleotides are present between these bases"
FT		13049. .13050
FT	misc_feature	/*tag= o
FT		/note= "approximately 180 unspecified nucleotides are present between these bases"
FT		13274. .13323
FT	misc_feature	/*tag= p
FT		/note= "these nucleotides as stated as being from AAV99111 in the figure, but are actually from AAV99112"
FT		13321. .13322
FT	misc_feature	/*tag= q
FT		/note= "approximately 1500 unspecified nucleotides are present between these bases"
FT		13328
FT	misc_feature	/*tag= r
FT		/note= "this nucleotide is not present in AAV99112"
FT		13566. .13567
FT	misc_feature	/*tag= s
FT		/note= "approximately 150 unspecified nucleotides are present between these bases"
FT		13827. .13828
FT	misc_feature	/*tag= t
FT		/note= "400 unspecified nucleotides are present between these bases"
FT		14073. .14074
FT	misc_feature	/*tag= u
FT		/note= "approximately 850 unspecified nucleotides are present between these bases"
FT		14359. .14360
FT	misc_feature	/*tag= v
FT		/note= "approximately 450 unspecified nucleotides are present between these bases"
FT		14796. .14797
FT	misc_feature	/*tag= w
FT		/note= "approximately 520 unspecified nucleotides are present between these bases"
FT		15194. .15195
FT	misc_feature	/*tag= x
FT		/note= "approximately 950 unspecified nucleotides are present between these bases"
FT		15869. .15870
FT	misc_feature	/*tag= y
FT		/note= "approximately 500 unspecified nucleotides are present between these bases"
FT		16324. .16325
FT	misc_feature	/*tag= z
FT		/note= "approximately 250 unspecified nucleotides are present between these bases"
FT		16764. .16765
FT	misc_feature	/*tag= aa
FT		/note= "approximately 1200 unspecified nucleotides are present between these bases"
FT		17085. .17086
FT	misc_feature	/*tag= ab
FT		/note= "approximately 600 unspecified nucleotides are present between these bases"
FT		17578. .17579
FT	misc_feature	/*tag= ac
FT		/note= "approximately 600 unspecified nucleotides are present between these bases"
FT		17868. .17869
FT	misc_feature	/*tag= ad
FT		/note= "approximately 800 unspecified nucleotides are present between these bases"

FT	misc_feature	18254..18253
FT		/tag= ae
FT		/note= "approximately 400 unspecified nucleotides are present between these bases"
FT	misc_feature	18611..18612
FT		/tag= af
FT		/note= "approximately 1450 unspecified nucleotides are present between these bases"
FT	misc_feature	18869..18870
FT		/tag= ag
FT		/note= "approximately 550 unspecified nucleotides are present between these bases"
XX		
FN	WO9854313-A2.	
XX		
PD	03-DEC-1998.	
XX		
PF	29-MAY-1998;	98WO-TB001107.
XX		
PR	30-MAY-1997;	97US-00866340.
PR	17-DEC-1997;	97US-0069865P.
XX		
FA	(TYMC-) UNIV MCGILL.	
XX		
PI	Szyf M, Bigey P, Ramchandani S;	
XX		
DR	WPI; 1999-059833/05.	
XX		
PT	New DNA methyltransferase nucleotide sequences - used particularly to develop antisense oligonucleotides for diagnostic and therapeutic purposes, particularly for inhibiting tumour growth.	
PS	Claim 1; Fig 1; 108pp; English.	
CC	The present sequence represents a full length DNA methyltransferase (DNMTase) genomic sequence. It is derived from fragments AAV99091-128. The sequences can be used to derive antisense oligonucleotides which inhibit DNMTase expression. The antisense oligonucleotides can be used in investigating the role of DNMTase in cellular growth. They can be administered at different points in the cell cycle, or in conjunction with promoters or inhibitors of cell growth to determine the role of DNMTase in the growth of the cell type of interest. The antisense oligonucleotides can also be used for inhibiting tumour growth in a mammal, or to activate silenced genes to provide a missing gene function. This ameliorates disease symptoms, e.g. in beta thalassemia and sickle cell anemia. The antisense oligonucleotides can also be used as analytical and diagnostic tools and potentiators of transgenic plant and animal studies	
SEQ	Sequence 19440 BP; 4471 A; 4929 C; 4936 G; 4915 T; 0 U; 189 Other;	
	Query Match	3.8%; Score 19; DB 2; Length 19440;
	Best Local Similarity	100.0%; Pred.No.45;
	Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	27 CATCCTGCTGAAGCTGGTC 45       	
Dd	16076 CATCCTGCTGAAGCTGGTC 16094       	
RESULT 42		
AAD54645		
ID	AAD54645 standard; DNA; 113193 BP.	
XX		
AC	AAD54645;	
XX		
DT	26-JUN-2003 (first entry)	
XX		
DE	Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.	
XX		
KW	Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.	
XX		
OS	Streptomyces nodosus.	

```

XX FH Location/Qualifiers
XX FT complement(4..1824)
XX FT /*tag= a
XX FT /product= "ABC transporter encoded by S. nodosus ampbG
XX FT gene"
XX FT complement(1805..3628)
XX FT /*tag= b
XX FT /product= "ABC transporter encoded by S. nodosus ampbH
XX FT gene"
XX FT 3840..4874
XX FT /*tag= c
XX FT /product= "GDP-mannose dehydratase encoded by S. nodosus
XX FT ampbDII gene"
XX FT 5042..33574
XX FT /*tag= d
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
XX FT nodosus amphi gene"
XX FT 33584..50518
XX FT /*tag= e
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 15, 16 and 17 encoded by S. nodosus
XX FT amphi gene"
XX FT 50571..56675
XX FT /*tag= f
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 18 and thioesterase encoded by S.
XX FT nodosus ampk gene"
XX FT 56829..58019
XX FT /*tag= g
XX FT /product= "Cytochrome P450 encoded by S. nodosus amphi
XX FT gene"
XX FT 58139..58648
XX FT /*tag= h
XX FT /product= "ORF1, hypothetical protein"
XX FT complement(58756..59610)
XX FT /*tag= i
XX FT /product= "ORF2, hypothetical protein"
XX FT 59869..61470
XX FT /*tag= j
XX FT /product= "Cytochrome P450 encoded by S. nodosus amphi
XX FT gene"
XX FT 61470..61995
XX FT /*tag= k
XX FT /product= "Ferredoxin encoded by S. nodosus ampbM gene"
XX FT complement(62051..63250)
XX FT /*tag= l
XX FT /product= "Cytochrome P450 encoded by S. nodosus amphi
XX FT gene"
XX FT 63250..64308
XX FT /*tag= m
XX FT /product= "NDP-sugar aminotransferase encoded by S.
XX FT nodosus ampbDII gene"
XX FT complement(64324..65775)
XX FT /*tag= n
XX FT /product= "Glycosyl transferase encoded by S. nodosus
XX FT ampbDI gene"
XX FT /transl_except= (pos:65773..65775, aa:Met)
XX FT 66081..70319
XX FT /*tag= o
XX FT /product= "Polyketide synthase multienzyme housing
XX FT loading module encoded by S. nodosus amphi gene"
XX FT 70366..79938
XX FT /*tag= p
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 1 and 2 encoded by S. nodosus by ampbH
XX FT gene"
XX FT 79956..112709
XX FT /*tag= q
XX FT /product= "Polyketide synthase multienzyme housing
XX FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
XX FT nodosus by ampk gene"

```

```

XX PN WO200297082-A2.
XX PD 05-DEC-2002.
XX PF 27-MAY-2002; 2002WO-IE000071.
XX PF 31-MAY-2001; 2001IE-00000527.
XX PR (UYDU-) UNIV COLLEGE DUBLIN.
XX PA Caffrey JP;
XX PI WPI; 2003-201271/19.
XX DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX DR AAE36129, AAE36130, AAE36131, AAE36132.
XX FT Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX FT for preparing amphotericin derivative or analog antibiotic agent with
XX FT altered properties, in biosynthesis of polyketide other than
XX FT amphotericin.
XX PS Claim 1; Page 52-114; 276pp; English.
XX CC The invention relates to the gene cluster encoding the polypeptides
XX CC responsible for the biosynthesis of the polyene antibiotic amphotericin
XX CC (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX CC useful for preparing amphotericin derivatives or analogue antibiotic
XX CC agents with altered properties and in the biosynthesis of polyketides
XX CC other than amphotericin. ampbDII, ampbDII or amphi mutants are useful
XX CC for producing amphotericin derivatives glycosylated with alternative
XX CC sugars; ampbDII or ampbDII gene sequences are useful in engineered
XX CC biosynthesis of perosaminyl-amphoteronolide B; ampbDII or ampbDII and
XX CC amphi gene sequences are useful in the engineered biosynthesis of
XX CC perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; ampbDII, ampbDII
XX CC and amphi gene sequences are useful for preparing polypeptides capable
XX CC of addition of mycosamine to a polyketide other than amphoteronolide A or
XX CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX CC The present sequence is S. nodosus amphi biosynthetic gene cluster
XX SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
Query Match 3.8%; Score 19; DB 7; Length 113193;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 121 GGCTTCGACTCGCTGGCCG 139
Db 69945 GGCTTCGACTCGCTGGCCG 69963
RESULT 43
AAI64291/c
ID AAI64291 standard; DNA; 128139 BP.
XX AC AAI64291;
XX DT 22-APR-2002 (first entry)
XX DE RRV genome nucleotide sequence.
XX KW RRV; rhesus rhadinovirus; Japanese macaque virus; multiple sclerosis;
XX KW JMVH; cytostatic; antiasthmatic; antiallergic; dermatological; vulnery;
XX KW gene therapy; leucopenia; thrombocytopaenia; inflammatory disease;
XX KW asthma; allergy; dermatitis; virus; ds.
XX OS Macaca mulatta rhadinovirus 17577.
XX PH Key Location/Qualifiers
XX FT 1353..2674
XX FT /*tag= a
XX FT /product= "RRV RI"

```

FT	CDS	complement(2692..3258)	FT	CDS	/label= RVV_ORF20
FT		/tag= b	FT		/note= "has similarity to KSHV ORF20"
FT		/product= "dihydrofolate reductase"	FT		32094..33767
FT		/label= RVV_ORF2	FT		/tag= f
FT		/note= "has similarity to Kaposi's sarcoma-associated	FT		/product= "thymidine kinase"
FT		virus (KSHV) open reading frame (ORF) 2"	FT		/label= RVV_ORF21
FT		3676..5613	FT		/note= "has similarity to KSHV ORF21"
FT		/tag= c	FT		33754..35868
FT		/product= "complement binding protein"	FT		/tag= s
FT		/label= RVV_ORF4	FT		/product= "glycoprotein H"
FT		/note= "has similarity to KSHV ORF4"	FT		/label= RVV_ORF22
FT		6045..9443	FT		/note= "has similarity to KSHV ORF22"
FT		/tag= d	FT		complement(35865..37073)
FT		/product= "ssDNA binding protein"	FT		/tag= t
FT		/label= RVV_ORF6	FT		/label= RVV_ORF23
FT		/note= "has similarity to KSHV ORF6"	FT		/note= "has similarity to KSHV ORF23"
FT		9468..11528	FT		complement(37123..39321)
FT		/tag= e	FT		/tag= u
FT		/product= "transport protein"	FT		/label= RVV_ORF24
FT		/label= RVV_ORF7	FT		/note= "has similarity to KSHV ORF24"
FT		/note= "has similarity to KSHV ORF7"	FT		39323..43459
FT		11515..14004	FT		/tag= v
FT		/tag= f	FT		/product= "major capsid protein"
FT		/product= "glycoprotein B"	FT		/label= RVV_ORF25
FT		/label= RVV_ORF8	FT		/note= "has similarity to KSHV ORF25"
FT		/note= "has similarity to KSHV ORF8"	FT		43491..44408
FT		14122..17166	FT		/tag= w
FT		/tag= g	FT		/product= "capsid protein"
FT		/product= "DNA polymerase protein"	FT		/label= RVV_ORF26
FT		/label= RVV_ORF9	FT		/note= "has similarity to KSHV ORF26"
FT		/note= "has similarity to KSHV ORF9"	FT		44433..45242
FT		17261..18511	FT		/tag= x
FT		/tag= h	FT		/label= RVV_ORF27
FT		/label= RVV_ORF10	FT		/note= "has similarity to KSHV ORF27"
FT		/note= "has similarity to KSHV ORF10"	FT		45408..45683
FT		18520..19749	FT		/tag= y
FT		/tag= i	FT		/label= RVV_ORF28
FT		/label= RVV_ORF11	FT		/note= "has similarity to KSHV ORF28"
FT		/note= "has similarity to KSHV ORF11"	FT		complement(45733..46779)
FT		complement(19921..20544)	FT		/tag= z
FT		/tag= j	FT		/label= RVV_ORF29b
FT		/product= "RVV R2"	FT		/note= "has similarity to KSHV ORF29b"
FT		/label= RVV_ORF	FT		46905..47135
FT		/note= "has similarity to KSHV interluekin (IL)-6 gene"	FT		/tag= aa
FT		complement(20777..21778)	FT		/label= RVV_ORF30
FT		/tag= k	FT		/note= "has similarity to KSHV ORF30"
FT		/product= "thymidylate synthase"	FT		47093..47746
FT		/label= RVV_ORF70	FT		/tag= ab
FT		/note= "has similarity to KSHV ORF70"	FT		/label= RVV_ORF31
FT		complement(22245..22592)	FT		/note= "has similarity to KSHV ORF31"
FT		/tag= l	FT		47683..49077
FT		/product= "RVV R3"	FT		/tag= ac
FT		/note= "has similarity to KSHV K4 viral MIP gene"	FT		/label= RVV_ORF32
FT		26846..27409	FT		/note= "has similarity to KSHV ORF32"
FT		/tag= m	FT		49049..50059
FT		/product= "Bcl2-homologue"	FT		/tag= ad
FT		/label= RVV_ORF16	FT		/label= RVV_ORF33
FT		/note= "has similarity to KSHV ORF16"	FT		/note= "has similarity to KSHV ORF33"
FT		complement(27515..29125)	FT		complement(49977..50960)
FT		/tag= n	FT		/tag= ae
FT		/label= RVV_ORF17	FT		/label= RVV_ORF29a
FT		/note= "has similarity to KSHV ORF17"	FT		/note= "has similarity to KSHV ORF29a"
FT		28998..29897	FT		50959..51942
FT		/tag= o	FT		/tag= af
FT		/label= RVV_ORF18	FT		/label= RVV_ORF34
FT		/note= "has similarity to KSHV ORF18"	FT		/note= "has similarity to KSHV ORF34"
FT		complement(29905..31548)	FT		51923..52372
FT		/tag= p	FT		/tag= ag
FT		/product= "tegument protein"	FT		/label= RVV_ORF35
FT		/label= RVV_ORF19	FT		/note= "has similarity to KSHV ORF35"
FT		/note= "has similarity to KSHV ORF19"	FT		52278..53585
FT		complement(31043..32095)	FT		/tag= ah
FT		/tag= q	FT		/product= "kinase"

FT /label= RVV\_ORF36  
FT /note= "has similarity to KSHV ORF36"  
FT CDS 53566..55008  
FT \*tag= ai  
FT /product= "alkaline exonuclease"  
FT /label= RVV\_ORF37  
FT /note= "has similarity to KSHV ORF37"  
FT CDS 54963..55172  
FT \*tag= aj  
FT /label= RVV\_ORF38  
FT /note= "has similarity to KSHV ORF38"  
FT CDS complement(55255..56391)  
FT \*tag= ak  
FT /product= "glycoprotein M"  
FT /label= RVV\_ORF39  
FT /note= "has similarity to KSHV ORF39"  
FT CDS 56526..57932  
FT \*tag= al  
FT /product= "helicase/primase"  
FT /label= RVV\_ORF40  
FT /note= "has similarity to KSHV ORF40"  
FT CDS 57917..58528  
FT \*tag= am  
FT /product= "helicase/primase"  
FT /label= RVV\_ORF41  
FT /note= "has similarity to KSHV ORF41"  
FT CDS complement(58525..59343)  
FT \*tag= an  
FT /label= RVV\_ORF42  
FT /note= "has similarity to KSHV ORF42"  
FT CDS complement(59297..61027)  
FT \*tag= ao  
FT /product= "capsid protein"  
FT /label= RVV\_ORF43  
FT /note= "has similarity to KSHV ORF43"

Query Match 3.8%; Score 19; DB 6; Length 128139;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 TGATCGACGACCTGGACAT 461  
Db 53480 TGATCGACGACCTGGACAT 53462

RESULT 44  
AAC64754/c  
ID AAC64754 standard; DNA; 133719 BP.  
XX  
AC AAC64754;  
XX  
DT 28-FEB-2001 (first entry)  
XX  
DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.  
XX  
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;  
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;  
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;  
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;  
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;  
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia;  
KW ds.  
XX  
OS Macaca mulatta rhadinovirus 17577.  
XX  
PN WO200028040-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US026260.  
XX  
PR 06-NOV-1998; 98US-0107507P.  
PR 20-NOV-1998; 98US-0109409P.

XX (UYOR-) UNIV OREGON HEALTH SCI.  
XX PA Wong SW, Axthelm MK, Searles RP;  
XX PI WPI; 2000-376552/32.  
XX DR  
XX PT New rhesus rhadino virus for producing non-human primate model useful for  
XX testing potential treatments and efficacy of the candidate vaccine for  
XX conditions associated with RRV infection.  
XX PS Claim 2; Page 83-122; 141pp; English.  
XX  
CC The present invention describes a novel rhesus macaque rhadinovirus  
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the  
CC RRV genome sequence, and AAB53213 to AAB53204 represent the proteins  
CC encoded by the genome sequence. The present invention also specifically  
CC claims the individual open reading frame (ORF) nucleotide sequences from  
CC the genome which encode the individual proteins, but these sequences are  
CC not given. A non-human animal infected with RRV can be used for testing  
CC the efficacy of drug in the treatment of condition associated with  
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative  
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,  
CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by  
CC administering the drug to an immuno-compromised non-human primate  
CC preferably Rhesus macaque monkey obtained by as a result of infection by  
CC Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-  
CC human primate model for testing potential treatments for conditions  
CC associated with RRV infection. It is also useful for testing the efficacy  
CC of the candidate vaccine against RRV infection or conditions associated  
CC with its infection by administering the vaccine to the subject capable of  
CC infection with RRV, inoculating the subject with RRV and observing the  
CC effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213  
CC represent sequence used in the exemplification of the present invention  
XX  
SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 U; 0 Other;  
  
Query Match 3.8%; Score 19; DB 3; Length 133719;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 443 TGATCGACGACCTGGACAT 461  
Db 59060 TGATCGACGACCTGGACAT 59042  
  
RESULT 45  
ACD93990  
ID ACD93990 standard; cDNA; 276 BP.  
XX  
AC ACD93990;  
XX  
DT 23-SEP-2003 (first entry)  
XX  
DE Human colon cancer cell expressed cDNA #2402.  
XX  
KW Open reading frame detection; genome sequencing; colon cancer;  
KW breast cancer; population genome analysis; genetic shift; cancer;  
KW antibiotic resistance; antibiotic non-tolerance; congenital disease;  
KW agriculture; food crop genome; resistance gene; retrovirus;  
KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;  
KW gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN US2002155438-A1.  
XX  
PD 24-OCT-2002.  
XX  
PF 27-SEP-1999; 99US-00406117.  
XX  
PR 20-NOV-1998; 98US-00196716.  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:48:58 ; Search time 57.3333 seconds  
(without alignments)  
4849.367 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 gcgggtcccgacgcgagca.....ccctcgccgacagactga 501

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 15

Total number of hits satisfying chosen parameters: 778

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgm2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgm2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgm2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgm2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgm2\_6/ptodata/2/ina/PCTUS COMB.seq.\*
- 6: /cgm2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	5.0	44377	2	US-08-804-227C-7
2	25	5.0	44377	2	US-08-804-198-1
C 3	23	4.6	748	3	US-09-154-083-1
C 4	23	4.6	753	3	US-09-154-083-28
5	23	4.6	4403765	3	US-09-103-840A-2
6	23	4.6	4411529	3	US-09-103-840A-1
7	21	4.2	47981	4	US-09-679-279-1
8	20	4.0	1410	4	US-09-252-991A-10147
C 9	20	4.0	1623	4	US-09-252-991A-9839
10	20	4.0	4026	4	US-07-642-734C-3
11	20	4.0	20235	1	US-07-642-734C-3
12	20	4.0	20235	3	US-08-439-009A-3
13	20	4.0	43280	2	US-08-804-227C-1
14	19	3.8	375	3	US-08-981-189B-16
15	19	3.8	451	3	US-08-866-340-24
16	19	3.8	452	3	US-09-103-875-30
17	19	3.8	953	4	US-09-452-239-45
18	19	3.8	1049	4	US-09-452-239-43
19	19	3.8	1368	1	US-08-578-590-1
C 20	19	3.8	1497	4	US-09-252-991A-2256
21	19	3.8	1950	4	US-09-252-991A-2425
C 22	18	3.6	579	4	US-09-252-991A-9199
23	18	3.6	717	4	US-09-252-991A-8972
24	18	3.6	819	4	US-09-252-991A-8861
25	18	3.6	1656	4	US-09-252-991A-8752
26	18	3.6	1791	1	US-08-399-646-3
27	18	3.6	1791	1	US-08-607-321-3

28	18	3.6	1791	2	US-08-961-240-3
29	18	3.6	1791	2	US-08-605-501-3
30	18	3.6	2056	1	US-08-393-646-13
31	18	3.6	2056	1	US-08-607-321-13
32	18	3.6	2056	2	US-08-961-240-13
33	18	3.6	2056	2	US-08-605-501-13
34	18	3.6	2508	4	US-09-252-991A-793
C 35	18	3.6	2577	4	US-09-252-991A-828
36	18	3.6	49377	1	US-08-764-233A-1
37	17	3.4	408	4	US-09-252-991A-2185
38	17	3.4	411	4	US-09-252-991A-8838
39	17	3.4	477	4	US-09-252-991A-12519
40	17	3.4	546	4	US-09-252-991A-5395
41	17	3.4	546	4	US-09-252-991A-12682
C 42	17	3.4	561	3	US-09-154-083-13
C 43	17	3.4	617	3	US-09-146-221-9
44	17	3.4	723	4	US-09-252-991A-5281
45	17	3.4	744	4	US-09-252-991A-5345

## ALIGNMENTS

RESULT 1  
US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-227C-7

Query Match      5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGCGCGCGGT 143
Db 35710 TCGGCTTCGACTCGCTGCGCGCGGT 35734

RESULT 2
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-198-1

Query Match      5.0%; Score 25; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGCGCGCGGT 143
Db 35710 TCGGCTTCGACTCGCTGCGCGCGGT 35734

RESULT 3
US-09-154-083-1/c
; Sequence 1, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 748
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-1

Query Match      4.6%; Score 23; DB 3; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGGCTGCGGCTGCCCGC 191
Db 366 GCCACCGGGCTGCGGCTGCCCGC 344

RESULT 4
US-09-154-083-28/c
; Sequence 28, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; TITLE OF INVENTION: Constructs Therefor
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-28

Query Match      4.6%; Score 23; DB 3; Length 753;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GCCACCGGGCTGCGGCTGCCCGC 191
Db 371 GCCACCGGGCTGCGGCTGCCCGC 349

RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
```

APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
Query Match 4.6%; Score 23; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 0.051;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 193 ACGCTGATCTTCGACTACCCAC 215  
Db 1872270 ACGCTGATCTTCGACTACCCAC 1872292  
RESULT 6  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Query Match 4.6%; Score 23; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.051;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 193 ACGCTGATCTTCGACTACCCAC 215  
Db 1881395 ACGCTGATCTTCGACTACCCAC 1881417  
RESULT 7  
US-09-679-279-1  
Sequence 1, Application US/09679279  
Patent No. 6524841  
GENERAL INFORMATION:  
APPLICANT: McDaniel, Robert  
APPLICANT: Volchegursky, Yanina  
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
TITLE OF INVENTION: Genes and Uses Thereof  
FILE REFERENCE: 300622004700  
CURRENT APPLICATION NUMBER: US/09/679,279  
CURRENT FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/158,305

PRIOR FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 60/190,024  
PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 47981  
TYPE: DNA  
ORGANISM: Micromonospora megalomicea  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(144)  
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (928)...(2061)  
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (2072)...(3382)  
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),  
OTHER INFORMATION: TDP-megosamine glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (3462)...(4634)  
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase  
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (4651)...(5775)  
OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DmJ homolog),  
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (5822)...(6595)  
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (6592)...(7197)  
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmU hc  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (7220)...(8206)  
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmV homolog),  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (8228)...(9220)  
OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (9226)...(10479)  
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (10483)...(11424)  
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (12181)...(22821)  
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (12505)...(13470)  
OTHER INFORMATION: megAL, AT-L  
NAME/KEY: misc feature  
LOCATION: (13576)...(13791)  
OTHER INFORMATION: megAI, ACP-L  
NAME/KEY: misc feature  
LOCATION: (13849)...(15126)  
OTHER INFORMATION: megAL, KS1  
NAME/KEY: misc feature



```

; LOCATION: (15427)...(16476)
; OTHER INFORMATION: megAI, AT1
; NAME/KEY: misc feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc feature
; LOCATION: (18268)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc feature
; LOCATION: (22319)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (32957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc feature
; LOCATION: (26998)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)

```

```

; OTHER INFORMATION: megAIII, AT6
; NAME/KEY: misc feature
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAIII, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAIII, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAIII, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megCIII, desosaminyl transferase, desosamine glycosyltransferase
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megBII-2 (megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ
; US-09-679-279-1

```

Query Match 4.2%; Score 21; DB 4; Length 47981;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 TTGACTCGCTGGCGCGGTC 144  
 ||||||||||||||||  
 Db 22447 TTGACTCGCTGGCGCGGTC 22467

RESULT 3  
 US-09-252-991A-10147  
 ; Sequence 10147, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 10147  
 ; LENGTH: 1410  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-10147

Query Match 4.0%; Score 20; DB 4; Length 1410;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGAGGCCGACGACGCGCTGG 277  
 ||||||||||||||||  
 Db 459 GGAGGCCGACGACGCGCTGG 478

RESULT 9  
 US-09-252-991A-9839/c

```
; Sequence 9839, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9839
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-9839

Query Match 4.0%; Score 20; DB 4; Length 1623;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGAGCCGCGACGCGCTGG 277
Db 170 GGAGCCGCGACGCGCTGG 151

RESULT 10
US-09-252-991A-10214
; Sequence 10214, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10214
; LENGTH: 4026
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-10214

Query Match 4.0%; Score 20; DB 4; Length 4026;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 GGAGCCGCGACGCGCTGG 277
Db 3746 GGAGCCGCGACGCGCTGG 3765

RESULT 11
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
```

```
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dancikers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722 /codon start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORE2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc feature
```

```

; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoacyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene =eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
; OTHER INFORMATION: 6-deoxyerythronolide B formatio"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximatr span of
; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 18379..18921

```

```

; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-07-642-734C-3
;
; Query Match 4.0%; Score 20; DB 1; Length 20235;
; Best Local Similarity 100.0%; Pred. No. 2.3;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 124 TTGACTCGCTGCGCGCGT 143
; Db 10354 TTGACTCGCTGCGCGCGT 10373
;
; RESULT 12
; US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NREL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19

```

OTHER INFORMATION: /function= "gene eryA"  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19..4470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3406..3921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..5847  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6054..7026  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 7165..9216  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: dehydratase and enoylreductase domains m"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9433..9984  
OTHER INFORMATION: /function= "approximate span  
OTHER INFORMATION: beta-ketoreductase of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10225..10483  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 4"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10723..20235  
OTHER INFORMATION: /codon start= 10723  
OTHER INFORMATION: /function= "gene =eryA"  
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6  
OTHER INFORMATION: 6-deoxyerythronolide B formatio"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10723..15165  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10831..12174

OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 12379..13350  
OTHER INFORMATION: /function= "approximatr span of  
OTHER INFORMATION: acyltransferase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14062..14610  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14857..15114  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15166..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15172..16569  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 16768..17721  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 18379..18921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19149..19398  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19492..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: thioesterase domain of module 6"  
US-08-439-009A-3  
Query Match 4.0%; Score 20; DB 3; Length 20235;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 124 TTCGACTCGCTGGCGCGGT 143  
Db 10354 TTCGACTCGCTGGCGCGGT 10373  
RESULT 13  
US-08-804-227C-1  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kunstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS

STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs  
TYPE: nucleic acid  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 816..14234  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14351..19945  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20010..31199  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31232..36067  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36249..41774  
US-08-804-227C-1

Query Match 4.0%; Score 20; DB 2; Length 43280;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 169 GCCACCGGGCTGGCGCTGCC 188  
Db 13893 GCCACCGGGCTGGCGCTGCC 13912

RESULT 14  
US-08-981-189B-16  
Sequence 16, Application US/08981189B  
Patent No. 6214797  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: UROCORTIN PEPTIDES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY  
STREET: 120 S. LaSalle Street, Suite 1600  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/981,189B

FILING DATE: 10-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,144  
FILING DATE: 13-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,223  
FILING DATE: 11-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Schumann, James J.  
REGISTRATION NUMBER: 20,856  
REFERENCE/DOCKET NUMBER: 57611  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-552-1311  
TELEFAX: 858-552-0095  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE: Human Genomic Placental Library  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..375  
US-08-981-189B-16

Query Match 3.8%; Score 19; DB 3; Length 375;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CGCGGAGCAGACCGCATC 30  
Db 333 CGCGGAGCAGACCGCATC 351

RESULT 15  
US-08-866-340-24  
Sequence 24, Application US/08866340  
Patent No. 6020318  
GENERAL INFORMATION:  
APPLICANT: Szyf, Moshe  
APPLICANT: Bigey, Pascal  
APPLICANT: Ramchandani, Shyam  
TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC  
SEQUENCES AND ANTISENSE OLIGONUCLEOTIDES  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States of America  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/866,340  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Keowl, Wayne A.  
REGISTRATION NUMBER: 33,923  
REFERENCE/DOCKET NUMBER: 106.101.187  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000

```

; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 451 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   US-08-866-340-24
;
Query Match          3.8%; Score 19; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45
Db 200 CATCTGCTGAAGCTGGTC 218

RESULT 16
US-09-103-875-30
; Sequence 30, Application US/09103875A
; Patent No. 6221849
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC SEQUENCES AND ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; FILE REFERENCE: 106101.194
; CURRENT APPLICATION NUMBER: US/09/103,875A
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: 60/069,865
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 08/866,340
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 30
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-103-875-30

Query Match          3.8%; Score 19; DB 3; Length 452;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 CATCTGCTGAAGCTGGTC 45
Db 201 CATCTGCTGAAGCTGGTC 219

RESULT 17
US-09-452-239-45
; Sequence 45, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyle-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1998-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 45
; LENGTH: 953
; TYPE: DNA
; ORGANISM: Triticum aestivum
; US-09-452-239-45

```

```

; ORGANISM: Triticum aestivum
; US-09-452-239-45
;
Query Match          3.8%; Score 19; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 CTGCTGAAGCTGGTCGCG 49
Db 591 CTGCTGAAGCTGGTCGCG 609

RESULT 18
US-09-452-239-43
; Sequence 43, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyle-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 43
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (948)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (992)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (994)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (999)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1003)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1038)
; US-09-452-239-43

Query Match          3.8%; Score 19; DB 4; Length 1049;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 CTGCTGAAGCTGGTCGCG 49
Db 591 CTGCTGAAGCTGGTCGCG 609

RESULT 19
US-08-578-590-1
; Sequence 1, Application US/08578590
; Patent No. 5817499
; GENERAL INFORMATION:
; APPLICANT: Dalboge, Henrik
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kofod, Lene V.
; US-08-578-590-1

```



```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8972
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8972

Query Match          3.6%; Score 18; DB 4; Length 717;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      410 CCGAAGCGGCCCGCGCG 427
Db      665 CCGAAGCGGCCCGCGCG 682

RESULT 24
US-09-252-991A-8861
; Sequence 8861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8861
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8861

Query Match          3.6%; Score 18; DB 4; Length 819;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      410 CCGAAGCGGCCCGCGCG 427
Db      549 CCGAAGCGGCCCGCGCG 566

RESULT 25
US-09-252-991A-8752
; Sequence 8752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8752
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

```

US-09-252-991A-8752

Query Match          3.6%; Score 18; DB 4; Length 1656;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      410 CCGAAGCGGCCCGCGCG 427
Db      97 CCGAAGCGGCCCGCGCG 114

RESULT 26
US-08-399-646-3
; Sequence 3, Application US/08399646
; Patent No. 5556781
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,646
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1791
US-08-399-646-3

Query Match          3.6%; Score 18; DB 1; Length 1791;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      123 CTTGACTCGCTGCGCG 140
Db      1044 CTTGACTCGCTGCGCG 1061

```



RESULT 27  
US-08-607-321-3  
; Sequence 3, Application US/08607321  
; Patent No. 5716813  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/607,321  
; FILING DATE: 26-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1791  
; US-08-607-321-3  
Query Match 3.6%; Score 18; DB 1; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 123 CTTGACTCGTGGCGGC 140  
Db 1044 CTTGACTCGTGGCGGC 1061  
RESULT 28  
US-08-961-240-3  
; Sequence 3, Application US/08961240  
; Patent No. 5830715  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji

; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,240  
; FILING DATE: 30-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,646  
; FILING DATE: 07-MAR-1995  
; APPLICATION NUMBER: JP 59834  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 59840  
; FILING DATE: 07-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA=5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1791  
; US-08-961-240-3  
Query Match 3.6%; Score 18; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 123 CTTGACTCGTGGCGGC 140  
Db 1044 CTTGACTCGTGGCGGC 1061  
RESULT 29  
US-08-605-501-3  
; Sequence 3, Application US/08605501  
; Patent No. 5834287  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: HATTORI, Kazuko  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,501  
FILING DATE: 26-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,646  
FILING DATE: 07-MAR-1995  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59840  
FILING DATE: 07-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA=5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1791  
US-08-605-501-3

Query Match 3.6%; Score 18; DB 2; Length 1791;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGC 140  
|||  
Db 1044 CTTGACTCGTGGCGC 1061

RESULT 30  
US-08-399-646-13  
Sequence 13, Application US/08399646  
Patent No. 5556781  
GENERAL INFORMATION:  
APPLICANT: KUBOTA, Michio  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: HATTORI, Kazuko  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/399,646  
FILING DATE: 07-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 59840  
FILING DATE: 07-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA=5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2056 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..1883  
US-08-399-646-13

Query Match 3.6%; Score 18; DB 1; Length 2056;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGACTCGTGGCGC 140  
|||  
Db 1136 CTTGACTCGTGGCGC 1153

RESULT 31  
US-08-607-321-13  
Sequence 13, Application US/08607321  
Patent No. 5716813  
GENERAL INFORMATION:  
APPLICANT: KUBOTA, Michio  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: HATTORI, Kazuko  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND  
TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/607,321  
FILING DATE: 26-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,646  
FILING DATE: 07-MAR-1995  
APPLICATION NUMBER: JP 59834  
FILING DATE: 07-MAR-1994

```

; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
;
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..1883
;
US-08-951-240-13
Query Match 3.6%; Score 18; DB 2; Length 2056;
Best Local Similarity 100.0%; Pred.No.25;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0
;
QY 123 CTTGCGACTCGCTGCCGCG 140
; |||||
Db 1136 CTTGCGACTCGCTGCCGCG 1153
;
RESULT 33
US-08-605-501-13
; Sequence 13, Application US/08605501
; Patent No. 5834287
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: HATTORI, Kazuko
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
; TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,501
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,646
; FILING DATE: 07-MAR-1995
; APPLICATION NUMBER: JP 59834
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 59840
; FILING DATE: 07-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA=5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 90..1883  
US-08-605-501-13

Query Match 3.6%; Score 18; DB 2; Length 2056;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 CTTGCGACTGCTGGCGC 140  
Db 1136 CTTGCGACTGCTGGCGC 1153

## RESULT 34

US-09-252-991A-793  
Sequence 793, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 793  
LENGTH: 2508  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-793

Query Match 3.6%; Score 18; DB 4; Length 2508;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GAGGCGCTGCTGGCTAC 237  
Db 1561 GAGGCGCTGCTGGCTAC 1578

## RESULT 35

US-09-252-991A-828/c  
Sequence 828, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 828  
LENGTH: 2577  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-828

Query Match 3.6%; Score 18; DB 4; Length 2577;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GAGGCGCTGCTGGCTAC 237  
Db 1800 GAGGCGCTGCTGGCTAC 1783

## RESULT 36

US-08-764-233A-1  
Sequence 1, Application US/08764233A  
Patent No. 5716849  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Schupp, Thomas  
APPLICANT: Beck, James J.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Neff, Snezana  
APPLICANT: Ryals, John A.  
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,233A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: 1506/CIP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Sorangium cellulosum  
IMMEDIATE SOURCE:  
CLONE: p98/1, pJL3, and pVKM15  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 927..19874  
OTHER INFORMATION: /product= "SorA"  
OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSs tha  
OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from  
OTHER INFORMATION: Saccharopolyspora erythraea."  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 927..19874  
OTHER INFORMATION: /product= "SorA"  
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs tha  
OTHER INFORMATION: are known to be involved in the synthesis of polyketide  
OTHER INFORMATION: compounds."  
FEATURE:

; NAME/KEY: misc feature  
; LOCATION: 942..7115  
; OTHER INFORMATION: /product= "Module 1 of SorA"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 7203..12884  
; OTHER INFORMATION: /product= "Module 2 of SorA"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 13455..19616  
; OTHER INFORMATION: /product= "Module 3 of SorA"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 19871..46318  
; OTHER INFORMATION: /product= "SorB"  
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 19870..24556  
; OTHER INFORMATION: /product= "Module 1 of SorB"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 24638..30920  
; OTHER INFORMATION: /product= "Module 2 of SorB"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 30881..35446  
; OTHER INFORMATION: /product= "Module 3 of SorB"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 35528..40114  
; OTHER INFORMATION: /product= "Module 4 of SorB"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 40190..46318  
; OTHER INFORMATION: /product= "Module 5 of SorB"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 46851..47891  
; OTHER INFORMATION: /product= "SorM"  
; OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."  
; OTHER INFORMATION: polyketide rapamycin."  
US-08-764-233A-1

Query Match 3.6%; Score 18; DB 1; Length 49377;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 CGCGCCACCGGCTGCG 182  
DB 13313 CGCGCCACCGGCTGCG 13330

## RESULT 37

US-09-252-991A-2185  
; Sequence 2185, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 2185  
; LENGTH: 408

; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-2185

Query Match 3.4%; Score 17; DB 4; Length 408;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 368 TCGCCGACACCGGCACC 384  
DB 361 TCGCCGACACCGGCACC 377

## RESULT 38

US-09-252-991A-8838  
; Sequence 8838, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8838  
; LENGTH: 411  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8838

Query Match 3.4%; Score 17; DB 4; Length 411;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 CGACGACGGCTGTGACG 280  
DB 197 CGACGACGGCTGTGACG 213

## RESULT 39

US-09-252-991A-12519  
; Sequence 12519, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12519  
; LENGTH: 477  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12519

Query Match 3.4%; Score 17; DB 4; Length 477;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCCGACACCGGCA 382  
DB 362 CCTCGCCGACACCGGCA 378

```

RESULT 40
US-09-252-991A-5395
; Sequence 5395, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5395
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5395

Query Match 3.4%; Score 17; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CGAGCAGAACCGCATCC 31
| | | | | | | | | | | | | | | | | | | | | |
Db 347 CGAGCAGAACCGCATCC 363

RESULT 41
US-09-252-991A-12682
; Sequence 12682, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12682
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12682

Query Match 3.4%; Score 17; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 CCTCGCCGACACCGGCA 382
| | | | | | | | | | | | | | | | | | | | | |
Db 127 CCTCGCCGACACCGGCA 143

RESULT 42
US-09-154-083-13/c
; Sequence 13, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz

```

```

; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-13

Query Match 3.4%; Score 17; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTG 135
| | | | | | | | | | | | | | | | | | | | | |
Db 295 TCGGCTTCGACTCGCTG 279

RESULT 43
US-09-146-221-9/c
; Sequence 9, Application US/09146221
; Patent No. 6180895
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; APPLICANT: Wildung, Mark R
; APPLICANT: McCaskill, David G
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el
; FILE REFERENCE: Transketolase, and Methods for the Expression Thereof
; CURRENT APPLICATION NUMBER: US/09/146,221
; CURRENT FILING DATE: 1998-09-01
; EARLIER APPLICATION NUMBER: 60/056,033
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 9
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Mentha piperita
US-09-146-221-9

Query Match 3.4%; Score 17; DB 3; Length 617;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CGGCGCCGAGGCGATCG 94
| | | | | | | | | | | | | | | | | | | | | |
Db 89 CGGCGCCGAGGCGATCG 73

RESULT 44
US-09-252-991A-5281
; Sequence 5281, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5281
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5281

```

```
Query Match      3.4%; Score 17; DB 4; Length 723;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      215 CCCCGGAGGCGCTGGTC 231
      |||
Db      513 CCCCGGAGGCGCTGGTC 529

RESULT 45
US-09-252-991A-5345
; Sequence 5345, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5345
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5345

Query Match      3.4%; Score 17; DB 4; Length 744;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      215 CCCCGGAGGCGCTGGTC 231
      |||
Db      544 CCCCGGAGGCGCTGGTC 560

Search completed: June 16, 2004, 19:32:22
Job time : 71.3333 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:56:30 ; Search time 241 Seconds  
(without alignments)

9488.873 Million cell updates/sec

Title: US-10-042-665a-3\_COPY\_15043\_15543

Perfect score: 501

Sequence: 1 cgggtcccgagcgagca.....ccctcgggcgagcgagctga 501

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2998549 seqs, 2282253817 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubna/US05\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	24	4.8	10056	15	US-10-156-761-411
2	24	4.8	86941	17	US-10-461-194-2
3	24	4.8	100000	15	US-10-156-761-15103
C 4	24	4.8	9025608	15	US-10-156-761-1
C 5	23	4.6	1552	13	US-10-425-114-1096
C 6	23	4.6	1583	13	US-10-425-114-962
C 7	23	4.6	14643	17	US-10-156-761-931
C 8	23	4.6	31422	15	US-10-204-862A-2
C 9	22	4.4	84428	17	US-10-229-148B-1
C 10	21	4.2	10892	15	US-10-156-761-414
C 11	20	4.0	1089	16	US-10-369-493-31550
C 12	20	4.0	2640	16	US-10-369-493-44289
C 13	20	4.0	4551	13	US-10-282-122A-30490
C 14	20	4.0	14520	15	US-10-156-761-2885

Sequence 930, App  
Sequence 2, Appli  
Sequence 15102, A  
Sequence 1, Appli  
Sequence 15, Appl  
Sequence 141, App  
Sequence 5550, Ap  
Sequence 45, Appl  
Sequence 538, Ap  
Sequence 7258, Ap  
Sequence 43, Appl  
Sequence 5634, Ap  
Sequence 8019, Ap  
Sequence 30672, A  
Sequence 10290, A  
Sequence 22, Appl  
Sequence 37369, A  
Sequence 14, Appl  
Sequence 285, App  
Sequence 172, App  
Sequence 381, App  
Sequence 1, Appli  
Sequence 3629, Ap  
Sequence 23847, A  
Sequence 154, App  
Sequence 42814, A  
Sequence 36804, A  
Sequence 9687, Ap  
Sequence 38849, A  
Sequence 39841, A  
Sequence 2391, Ap

US-10-156-761-930  
US-10-204-862A-2  
US-10-156-761-15102  
US-10-156-761-1  
US-09-818-009-16  
US-10-198-846-141  
US-10-156-761-5550  
US-09-452-239-45  
US-10-156-761-5938  
US-10-156-761-7258  
US-09-452-239-43  
US-10-156-761-5634  
US-09-815-242-8019  
US-10-282-122A-30672  
US-10-198-846-10290  
US-10-250-613-22  
US-10-369-493-37369  
US-10-450-826-14  
US-09-962-436-285  
US-09-954-531-172  
US-09-954-531-381  
US-10-144-577-1  
US-10-156-761-3629  
US-10-425-114-23847  
US-10-389-647-154  
US-10-369-493-42814  
US-10-282-122A-36804  
US-09-815-242-9687  
US-10-282-122A-38849  
US-10-282-122A-39841  
US-10-156-761-2391

# ALIGNMENTS

RESULT 1  
US-10-156-761-411  
; Sequence 411, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 411  
; LENGTH: 10056  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(10056)  
US-10-156-761-411

Query Match 4.8%; Score 24; DB 15; Length 10056;  
Best Local Similarity 100.0%; Pred. No. 0.027; 0;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 CTGCCGCGACGCTGATCTCGAC 207  
Db 4456 CTGCCGCGACGCTGATCTCGAC 4479



```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Query Match      4.8%; Score 24; DB 15; Length 100000;
Best Local Similarity 100.0%; Pred. No. 0.017; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

QY      184 CTGCCGCGGACGCTGATCTTCGAC 207
DB      73463 CTGCCGCGGACGCTGATCTTCGAC 73486

RESULT 4
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      4.8%; Score 24; DB 15; Length 9025608;
Best Local Similarity 100.0%; Pred. No. 0.0065; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

QY      184 CTGCCGCGGACGCTGATCTTCGAC 207
DB      504449 CTGCCGCGGACGCTGATCTTCGAC 504426

RESULT 5
US-10-425-114-1096/c
; Sequence 1096, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1096
; LENGTH: 1552
; TYPE: DNA
```

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700101596_FLI
US-10-425-114-1096

Query Match          4.6%; Score 23; DB 13; Length 1552;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGCGTGCC 320
Db 294 CGCGAGTCTCTCGCGCGCGTGCC 272

RESULT 6
US-10-425-114-962/c
; Sequence 962, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 962
; LENGTH: 1583
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700097375_FLI
US-10-425-114-962

Query Match          4.6%; Score 23; DB 13; Length 1583;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGCGTGCC 320
Db 320 CGCGAGTCTCTCGCGCGCGTGCC 298

RESULT 7
US-10-156-761-931
; Sequence 931, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 931
; LENGTH: 14643
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(14643)
US-10-156-761-931

Query Match          4.6%; Score 23; DB 15; Length 14643;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGCTGATCTTCGACTACCCAC 215
Db 4645 ACGCTGATCTTCGACTACCCAC 4667

RESULT 8
US-10-204-862A-2
; Sequence 2, Application US/10204862A
; Publication No. US20040101936A1
; GENERAL INFORMATION:
; APPLICANT: ENDO, HIROFUMI
; APPLICANT: YAMAGUCHI, HIROYUKI
; APPLICANT: KANDA, YUTAKA
; APPLICANT: HASHIMOTO, SHINICHI
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE
; FILE REFERENCE: 468-32
; CURRENT APPLICATION NUMBER: US/10/204,862A
; CURRENT FILING DATE: 2003-12-26
; PRIOR APPLICATION NUMBER: JP 00/047405
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 31422
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(14643)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14824)..(31419)
US-10-204-862A-2

Query Match          4.6%; Score 23; DB 17; Length 31422;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ACGCTGATCTTCGACTACCCAC 215
Db 4645 ACGCTGATCTTCGACTACCCAC 4667

RESULT 9
US-10-229-148B-1
; Sequence 1, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: Complement((1)..(675))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((113016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; NAME/KEY: CDS

; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
```

```
/ OTHER INFORMATION: ORF12
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: Complement((79391)..(81052))
/ OTHER INFORMATION: ORF13
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (81541)..(82356)
/ OTHER INFORMATION: ORF14
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (82760)..(83362)
/ OTHER INFORMATION: ORF15
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: Complement((83495)..(84142))
/ OTHER INFORMATION: ORF16
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (84329)..(84428)
/ OTHER INFORMATION: ORF17 (fragment)
US-10-229-148B-1

Query Match          4.4%; Score 22; DB 17; Length 84428;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 TCGGCTTCGACTCGCTGGCGC 140
Db 64243 TCGGCTTCGACTCGCTGGCGC 64264

RESULT 10
US-10-156-761-414
/ Sequence 414, Application US/10156761
/ Publication No. US2003019018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 414
/ LENGTH: 10692
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(10692)
US-10-156-761-414

Query Match          4.2%; Score 21; DB 15; Length 10692;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 CTGGGCTGCCCGAGCGTG 198
Db 10357 CTGGGCTGCCCGAGCGTG 10377

RESULT 11
US-10-369-493-31550/c

/ Sequence 31550, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 31550
/ LENGTH: 1089
/ TYPE: DNA
/ ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31550

Query Match          4.0%; Score 20; DB 16; Length 1089;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATCC 31
Db 208 CGCCGAGCAGACCGCATCC 189

RESULT 12
US-10-369-493-44289
/ Sequence 44289, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 44289
/ LENGTH: 2640
/ TYPE: DNA
/ ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44289

Query Match          4.0%; Score 20; DB 16; Length 2640;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 CGCGCGCGAAGGCATCGGCC 97
Db 1683 CGCGCGCGAAGGCATCGGCC 1702

RESULT 13
US-10-282-122A-30490
/ Sequence 30490, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
```

Thu Jun 17 11:20:19 2004

```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30490
; LENGTH: 4551
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30490

Query Match 4.0%; Score 20; DB 13; Length 4551;
Best Local Similarity 100.0%; Pred.No. 3.1; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 258 GGAGCGCGACGACGCGCTGG 277
Db 3600 GGAGCGCGACGACGCGCTGG 3619

RESULT 14
US-10-156-761-2885
; Sequence 2885, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2885
; LENGTH: 14520
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(16596)
US-10-156-761-2885

Query Match 4.0%; Score 20; DB 15; Length 14520;
Best Local Similarity 100.0%; Pred.No. 2.4; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 172 ACCGGGCTCGGCTGCCGCG 191
Db 4270 ACCGGGCTCGGCTGCCGCG 4289

RESULT 15
US-10-156-761-930/c
; Sequence 930, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 930
; LENGTH: 16596
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(16596)
US-10-156-761-930

Query Match 4.0%; Score 20; DB 15; Length 16596;
Best Local Similarity 100.0%; Pred.No. 2.4; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

QY 366 CCTCGCGACACCGGACCG 385
Db 12387 CCTCGCGACACCGGACCG 12368

RESULT 16
US-10-204-862A-2/c
; Sequence 2, Application US/10204862A
; Publication No. US20040101936A1
; GENERAL INFORMATION:
; APPLICANT: ENDO, HIROFUMI
; APPLICANT: YAMAGUCHI, HIROYUKI
; APPLICANT: KANDA, YUTAKA
; APPLICANT: HASHIMOTO, SHINICHI
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; TITLE OF INVENTION: PROCESS FOR PRODUCING AVERMECTIN DERIVATIVE
; FILE REFERENCE: 468-32
; CURRENT APPLICATION NUMBER: US/10/204,862A
; CURRENT FILING DATE: 2003-12-26
; PRIOR APPLICATION NUMBER: JP 00/047405
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 31422
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(14643)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14824)..(31419)
US-10-204-862A-2

Query Match          4.0%; Score 20; DB 17; Length 31422;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 CCTGCCGACACCGGCACCG 385
Db 27210 CCTGCCGACACCGGCACCG 27191

RESULT 17
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match          4.0%; Score 20; DB 15; Length 125746;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 ACCGGGCTGCGGCTGCCCGC 191
Db 48046 ACCGGGCTGCGGCTGCCCGC 48065

RESULT 18
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
```

```
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          4.0%; Score 20; DB 15; Length 9025608;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 CCTGCCGACACCGGCACCG 385
Db 1175384 CCTGCCGACACCGGCACCG 1175383

RESULT 19
US-09-818-009-16
; Sequence 16, Application US/09818009
; Publication No. US20030032587A1
; GENERAL INFORMATION:
; APPLICANT: THE SALK INSTITUTE FOR BIOLOGICAL STUDIES
; TITLE OF INVENTION: UROCORTIN PEPTIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY
; STREET: 120 S. LaSalle Street, Suite 1600
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/818,009
; APPLICATION NUMBER: US/09/818,009
; FILING DATE: 26-Mar-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/981,189
; FILING DATE: 10-DEC-1997
; APPLICATION NUMBER: US 60/028,144
; FILING DATE: 13-JUN-1995
; APPLICATION NUMBER: US 60/002,223
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Schumann, James J.
; REGISTRATION NUMBER: 20,856
; REFERENCE/DOCKET NUMBER: 57611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-552-1311
; TELEFAX: 858-552-0095
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human Genomic Placental Library
```

```
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..375
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-818-009-16
    Query Match          3.8%; Score 19; DB 10; Length 375;
    Best Local Similarity 100.0%; Pred. No. 17;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 333 CGCCGAGCAGAACCGCATC 351

RESULT 20
US-10-198-846-141/c
; Sequence 141, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 380..402
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-141
    Query Match          3.8%; Score 19; DB 15; Length 485;
    Best Local Similarity 100.0%; Pred. No. 16;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGAACCGCATC 30
Db 181 CGCCGAGCAGAACCGCATC 163

RESULT 21
US-10-156-761-5550/c
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
;
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5550
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-156-761-5550
    Query Match          3.8%; Score 19; DB 15; Length 726;
    Best Local Similarity 100.0%; Pred. No. 14;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GAAGACGACCTCCGGCGAG 304
Db 528 GAAGACGACCTCCGGCGAG 510

RESULT 22
US-09-452-239-45
; Sequence 45, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: 60/110,594
; PRIOR FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 45
; LENGTH: 953
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-45
    Query Match          3.8%; Score 19; DB 9; Length 953;
    Best Local Similarity 100.0%; Pred. No. 14;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGGTCCGCG 49
Db 591 CTGCTGAAGCTGGTCCGCG 609

RESULT 23
US-10-156-761-5938
; Sequence 5938, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5938
; LENGTH: 996
```

```
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(996)
US-10-156-761-5938

Query Match          3.8%; Score 19; DB 15; Length 996;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 CGCGCCACCGGCTGCGG 183
Db 30 CGCGCCACCGGCTGCGG 48

RESULT 24
US-10-156-761-7258/c
; Sequence 7258, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7258
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
US-10-156-761-7258

Query Match          3.8%; Score 19; DB 15; Length 1044;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 CGCTGGAAGCTGCTCGGGA 260
Db 809 GCGTGAAGCTGCTCGGGA 791

RESULT 25
US-09-452-239-43
; Sequence 43, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 43
; LENGTH: 1049

; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (352)
; NAME/KEY: unsure
; LOCATION: (948)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (992)
; NAME/KEY: unsure
; LOCATION: (994)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (999)
; NAME/KEY: unsure
; LOCATION: (1003)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1038)
US-09-452-239-43

Query Match          3.8%; Score 19; DB 9; Length 1049;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGCTCGCG 49
Db 591 CTGCTGAAGCTGCTCGCG 609

RESULT 26
US-10-156-761-5634
; Sequence 5634, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5634
; LENGTH: 1293
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1293)
US-10-156-761-5634

Query Match          3.8%; Score 19; DB 15; Length 1293;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 GCTGGACACGCTGCTCGGC 366
Db 99 GCTGGACACGCTGCTCGGC 117
```



```
RESULT 27
US-09-815-242-8019
; Sequence 8019, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8019
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1893)
US-09-815-242-8019
Query Match 3.8%; Score 19; DB 9; Length 1893;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 CGAAGGCATCGCCCGCGC 102
Db 807 CGAAGGCATCGCCCGCGC 825

RESULT 28
US-10-282-122A-30672
; Sequence 30672, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

US-10-042-665a-3_c_15043_15543.ol.rnpb
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30672
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30672
Query Match 3.8%; Score 19; DB 13; Length 1893;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 CGAAGGCATCGCCCGCGC 102
Db 807 CGAAGGCATCGCCCGCGC 825

RESULT 29
US-10-198-846-10290
; Sequence 10290, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10290
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 1829, 1830, 1831, 1997, 1998, 1999, 2000
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10290
Query Match 3.8%; Score 19; DB 15; Length 2000;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CGCCGAGCAGAACCGCATC 30
Db 807 CGAAGGCATCGCCCGCGC 825
```

```

Db      701 CGCCGAGCAGAACCGCATC 719

RESULT 30
US-10-613-22/c
; Sequence 22, Application US/10250613
; Publication No. US20040096828A1
; GENERAL INFORMATION:
; APPLICANT: LU, Dying Aina M.; BAUGHN, Mariah R.;
; APPLICANT: YAO, Monique G.; DING, Li;
; APPLICANT: HONCHELL, Cynthia D.; YUE, Henry;
; APPLICANT: TANG, Y. Tom; WARREN, Bridget A.;
; APPLICANT: DUGGAN, Brendan M.; XU, Yuming;
; APPLICANT: CHAWLA, Narinder K.; GRIFFIN, Jennifer A.;
; APPLICANT: STEWART, Elizabeth A.; GANDHI, Ameena R.;
; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
; APPLICANT: ISON, Craig H.; AZIMZAI, Yalda;
; APPLICANT: HAFALIA, April J.A.; GIETZEN, Kimberly J.;
; APPLICANT: LAL, Preeti G.; SANJANWALA, Madhusudan M.;
; APPLICANT: ELLIOTT, Vicki S.
; TITLE OF INVENTION: CYTOSKELETAL-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0878 USN
; CURRENT APPLICATION NUMBER: US/10/250,613
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: PCT/US02/00178
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,085
; PRIOR FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/268,554
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/269,111
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 60/271,211
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7131221CBI
US-10-250-613-22

Query Match      3.8%; Score 19; DB 17; Length 2040;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CGCCGAGCAGAACCGCATC 30
        |||||||
Db      1949 CGCCGAGCAGAACCGCATC 1931

RESULT 31
US-10-369-493-37369
; Sequence 37369, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37369

```

```

; LENGTH: 2856
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37369

Query Match      3.8%; Score 19; DB 16; Length 2856;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      79 GCGCGGAGGCGATCGGCC 97
        |||||||
Db      2035 GCGCGGAGGCGATCGGCC 2053

RESULT 32
US-10-450-826-14
; Sequence 14, Application US/10450826
; Publication No. US20040101818A1
; GENERAL INFORMATION:
; APPLICANT: Ji, Darren
; APPLICANT: Axelrod, Douglas W.
; APPLICANT: Cook, Jonathon S.
; APPLICANT: Jaiswal, Neelam
; APPLICANT: Bistein, Richard
; APPLICANT: Houghton, Adam
; APPLICANT: Mertz, Lawrence
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation
; FILE REFERENCE: 044921-5039-WO
; CURRENT APPLICATION NUMBER: US/10/450,826
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/255,882
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/285,691
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. AF038633
US-10-450-826-14

Query Match      3.8%; Score 19; DB 17; Length 3081;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 CGCCGAGCAGAACCGCATC 30
        |||||||
Db      2915 CGCCGAGCAGAACCGCATC 2933

RESULT 33
US-09-962-436-285
; Sequence 285, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 285
; LENGTH: 5434
; TYPE: DNA

```

```
; ORGANISM: Homo sapiens
US-09-962-436-285

Query Match      3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      27 CATCCTGCTGAAGCTGGTC 45
Db      3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 34
US-09-954-531-172
; Sequence 172, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 172
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-172

Query Match      3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      27 CATCCTGCTGAAGCTGGTC 45
Db      3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 35
US-09-954-531-381
; Sequence 381, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 381
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-381

Query Match      3.8%; Score 19; DB 9; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      27 CATCCTGCTGAAGCTGGTC 45
Db      3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 36
US-10-144-577-1
; Sequence 1, Application US/10144577
; Publication No. US20030083292A1
; GENERAL INFORMATION:
; APPLICANT: MacLeod, Alan Robert
; TITLE OF INVENTION: Inhibitors of DNA Methyltransferase Isoforms
; FILE REFERENCE: MET-005
; CURRENT APPLICATION NUMBER: US/10/144,577
; CURRENT FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/290,202
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,212
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-144-577-1

Query Match      3.8%; Score 19; DB 15; Length 5434;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      27 CATCCTGCTGAAGCTGGTC 45
Db      3813 CATCCTGCTGAAGCTGGTC 3831

RESULT 37
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
; LENGTH: 11058
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
```

; LOCATION: (1)..(11058)  
 US-10-156-761-3629

Query Match 3.8%; Score 19; DB 15; Length 11058;  
 Best Local Similarity 100.0%; Pred. No. 8.2;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 GCTCGGCGCTGGCGACAC 378  
 |||||  
 DB 10500 GCTCGGCGCTGGCGACAC 10518

RESULT 38  
 US-10-425-114-23847  
 ; Sequence 23847, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 23847  
 ; LENGTH: 520  
 ; TYPE: DNA  
 ; ORGANISM: Chlorella sorokiniana  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3602-017-G8\_FLI  
 US-10-425-114-23847

Query Match 3.6%; Score 18; DB 13; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 GAGCGGGGCTGCTGCAC 354  
 |||||  
 DB 30 GAGCGGGGCTGCTGCAC 47

RESULT 39  
 US-10-389-647-154  
 ; Sequence 154, Application US/10389647  
 ; Publication No. US20040033549A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GREENBERG, E. Peter  
 ; APPLICANT: SCHUSTER, Martin  
 ; APPLICANT: LOSTROH, Candi  
 ; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA  
 ; FILE REFERENCE: UIZ-038CP  
 ; CURRENT APPLICATION NUMBER: US/10/389,647  
 ; CURRENT FILING DATE: 2003-03-14  
 ; PRIOR APPLICATION NUMBER: 09/653730  
 ; PRIOR FILING DATE: 2000-09-01  
 ; PRIOR APPLICATION NUMBER: 60/153022  
 ; PRIOR FILING DATE: 1999-09-03  
 ; NUMBER OF SEQ ID NOS: 710  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 154  
 ; LENGTH: 546  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-10-389-647-154

Query Match 3.6%; Score 18; DB 13; Length 546;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 CCGAAGCGGCGCGCGCG 427  
 |||||  
 DB 494 CCGAAGCGGCGCGCGCG 511

RESULT 40  
 US-10-369-493-42814  
 ; Sequence 42814, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 42814  
 ; LENGTH: 855  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-10-369-493-42814

Query Match 3.6%; Score 18; DB 16; Length 855;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 GCGGGGCTGCTGCAC 356  
 |||||  
 DB 432 GCGGGGCTGCTGCAC 449

RESULT 41  
 US-10-282-122A-36804  
 ; Sequence 36804, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA 034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 36804  
; LENGTH: 960  
; TYPE: DNA  
; ORGANISM: Salmonella paratyphi A  
US-10-282-122A-36804

Query Match 3.6%; Score 18; DB 13; Length 960;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCCACA 76  
|||||  
Db 734 CGACGGTGTCTCGGCCACA 751

RESULT 42  
US-09-815-242-9687  
; Sequence 9687, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; Prokaryotes  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9687  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(963)  
US-09-815-242-9687

Query Match 3.6%; Score 18; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCCACA 76

Db 734 CGACGGTGTCTCGGCCACA 751  
|||||  
RESULT 43  
US-10-282-122A-38849  
; Sequence 38849, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 38849  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhimurium

US-10-282-122A-38849  
Query Match 3.6%; Score 18; DB 13; Length 963;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CGACGGTGTCTCGGCCACA 76  
|||||  
Db 734 CGACGGTGTCTCGGCCACA 751

RESULT 44  
US-10-282-122A-39841  
; Sequence 39841, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39841  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Salmonella typhi  
US-10-282-122A-39841

Query Match 3.6%; Score 18; DB 13; Length 963;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 CGACGGTGCTCGGCACA 76  
|||||  
Db 734 CGACGGTGCTCGGCACA 751

RESULT 45  
US-10-156-761-2391  
; Sequence 2391, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2391  
; LENGTH: 1245  
; TYPE: DNA

; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1245)  
US-10-156-761-2391  
  
Query Match 3.6%; Score 18; DB 15; Length 1245;  
Best Local Similarity 100.0%; Pred.No. 41;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 356 CGCTGCTCGGCTCGCCG 373  
|||||  
Db 374 CGCTGCTCGGCTCGCCG 391  
  
Search completed: June 16, 2004, 20:02:40  
Job time : 1336 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2004, 15:37:20 ; Search time 2102.33 Seconds  
(without alignments)  
7116.356 Million cell updates/sec

Title: US-10-042-665A-3\_COPY\_15043\_15543  
Perfect score: 501  
Sequence: 1 gcggtcccgacgcgagca.....ccctcggcagacgagctga 501

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 15  
Total number of hits satisfying chosen parameters: 16108

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_nam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gasl:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	4.6	479	14	CF650732
C 2	23	4.6	501	12	B1273373
C 3	23	4.6	601	9	AI973519
C 4	23	4.6	742	29	CG384197

C 5	23	4.6	820	29	CG297728
C 6	23	4.6	829	28	BZ731737
C 7	23	4.6	857	28	CG381148
C 8	23	4.6	858	28	BZ731730
C 9	23	4.6	872	29	CG092377
C 10	23	4.6	876	29	CG289214
C 11	23	4.6	900	28	CC381145
C 12	23	4.6	918	29	CG384209
C 13	23	4.6	974	29	CG289224
C 14	23	4.6	1889	11	AY103974
C 15	22	4.4	340	14	DI5897
C 16	21	4.2	629	14	CA741791
C 17	21	4.2	1189	10	AW982464
C 18	20	4.0	354	14	CF770975
C 19	20	4.0	375	14	CF759576
C 20	20	4.0	491	12	BM371243
C 21	20	4.0	683	12	BI957736
C 22	20	4.0	702	28	CC346318
C 23	20	4.0	717	12	BG962173
C 24	20	4.0	770	29	CG670138
C 25	20	4.0	783	29	CG382731
C 26	20	4.0	802	29	CG382740
C 27	20	4.0	849	28	BZ649718
C 28	20	4.0	1035	10	BE791387
C 29	19	3.8	112	14	CB927303
C 30	19	3.8	211	10	BF822655
C 31	19	3.8	219	9	AI352333
C 32	19	3.8	219	9	AI631150
C 33	19	3.8	220	10	BF773138
C 34	19	3.8	251	9	AI971801
C 35	19	3.8	346	9	AI716696
C 36	19	3.8	349	10	BE415413
C 37	19	3.8	356	10	BF941825
C 38	19	3.8	358	14	CA646451
C 39	19	3.8	379	14	H24268
C 40	19	3.8	384	10	BF523813
C 41	19	3.8	399	14	CA614966
C 42	19	3.8	410	10	BE415414
C 43	19	3.8	414	10	BF474477
C 44	19	3.8	439	10	AW294748
C 45	19	3.8	453	12	EM358573

ALIGNMENTS

RESULT 1  
CF650732/c  
LOCUS  
DEFINITION  
3530\_1\_92\_1\_E03.x.1\_3530 - Full length cDNA library created by  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CF650732 479 bp mRNA linear EST 02-OCT-2003  
3530\_1\_92\_1\_E03.x.1\_3530 - Full length cDNA library created by  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.  
CF650732  
CF650732.1 GI:37425965  
EST.  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 479)  
Walbot.V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 3530\_1\_92\_1 row: E column: 03.  
Location/Qualifiers

```

source
1. .479
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="multiple"
/dev_stage="varies by tissue"
/lab_host="DH10B"
/clone_lib="3530 - Full length cDNA library created by
Invitrogen from multiple tissues"
/notes="Organ: silks, husks, ears, pollen, shoot tips,
leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT
6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery
Project contracted with Invitrogen to produce a
normalized, full length library in a pSPORT vector. This
is a Gateway compatible vector, permitting clone movement
to new vector backbones for expression in diverse host
cells using recombination rather than restriction enzymes.
Details of the vector and sequencing primers are available
at ZmDB in the EST library description tables. poly(A)+
mRNA was prepared by Invitrogen, and equimolar amounts of
RNA from each of the 12 tissue samples were mixed together
for selection of mRNA with a 5' cap. After synthesis of
cDNA, a normalization step was conducted against the
mixture of RNA sources. This step effected a 20X to 80X
reduction in common transcript types. Tissues prepared: 1.
just emerging silks; 2. inner husks from ears of sample
#1; 3. 20 dap aleurone; 4. immature tassels, stages from
1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm
vegetative shoot tips from 15 day old seedlings; all
leaves with an expanded or partially expanded sheath
were removed; 8. mature leaf tissue; 9. 0.5 cm long root
tips from 15 day old seedlings; 10. 10 dap whole seed; 11.
12 dap endosperm and embryo; 12. 17 dap endosperm and
embryo. All of the sequenced clones in project 3530 will
be archived at the University of Arizona along with the
unigene clones from the Maize Gene Discovery EST
sequencing projects. Clones can be ordered through the
ZmDB web site or directly from the University of Arizona
(http://www.genome.arizona.edu/orders/). High density
filters containing over 18,000 clones can also be ordered
from the University of Arizona."

ORIGIN
Query Match 4.6%; Score 23; DB 14; Length 479;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCCTCGCGCGCGTGC 320
| | | | | | | | | | | | | | | | | | | | |
Db 354 CGCGAGTCCTCGCGCGCGTGC 332
| | | | | | | | | | | | | | | | | | | | |

RESULT 2
B1273373/c
LOCUS
DEFINITION B1273373 501 bp mRNA linear EST 18-JUL-2001
Zea mays cDNA, mRNA sequence.
ACCESSION B1273373
VERSION B1273373.1 GI:14883487
KEYWORDS EST.
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 501)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

source
1. .501
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W64A"
/db_xref="taxon:4577"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/notes="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

ORIGIN
Query Match 4.6%; Score 23; DB 12; Length 501;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCCTCGCGCGCGTGC 320
| | | | | | | | | | | | | | | | | | | | |
Db 319 CGCGAGTCCTCGCGCGCGTGC 297
| | | | | | | | | | | | | | | | | | | | |

RESULT 3
A1973519/c
LOCUS
DEFINITION A1973519 601 bp mRNA linear EST 25-AUG-1999
496027F10.xl 496 - stressed shoot cDNA library from Wang/Bohnert
lab Zea mays cDNA, mRNA sequence.
ACCESSION A1973519
VERSION A1973519.1 GI:5770345
KEYWORDS EST.
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 601)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

```

```

Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496024 row: B column: 09.
Location/Qualifiers
1. .501
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W64A"
/db_xref="taxon:4577"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/notes="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

ORIGIN
Query Match 4.6%; Score 23; DB 12; Length 501;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCCTCGCGCGCGTGC 320
| | | | | | | | | | | | | | | | | | | | |
Db 319 CGCGAGTCCTCGCGCGCGTGC 297
| | | | | | | | | | | | | | | | | | | | |

RESULT 3
A1973519/c
LOCUS
DEFINITION A1973519 601 bp mRNA linear EST 25-AUG-1999
496027F10.xl 496 - stressed shoot cDNA library from Wang/Bohnert
lab Zea mays cDNA, mRNA sequence.
ACCESSION A1973519
VERSION A1973519.1 GI:5770345
KEYWORDS EST.
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 601)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496027 row: F column: 10.
Location/Qualifiers
1. .601
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"

```



```

/db_xref="taxon:4577"
/tissue_type="seedling"
/dev_stages="salt stress"
/lab_host="E.coli XL Gold"
/clone_lib="496 - stressed shoot cDNA library from
Wang/Bohnert lab"
/note="Organ: shoot; Vector: pBluescriptII SK(+) XR;
Wang/Bohnert"

ORIGIN
Query Match 4.6%; Score 23; DB 9; Length 601;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGCC 320
Db 288 CGCGAGTCTCGCGCGCGTGCC 266

RESULT 4
CG384197
LOCUS CG384197 742 bp DNA linear GSS 26-AUG-2003
DEFINITION OG08R89TH ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0681P10,
genomic survey sequence.
ACCESSION CG384197
VERSION CG384197.1 GI:34301464
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 742)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OG08R89TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..742
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBma0681P10"
/clone_lib="ZM.0.7.1.5 KB"
/note="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 29; Length 820;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGCC 320
Db 506 CGCGAGTCTCGCGCGCGTGCC 484

RESULT 6
BZ731737/c
LOCUS BZ731737/c 829 bp DNA linear GSS 03-MAR-2003
DEFINITION OGFAS22TM ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0241P09,
genomic survey sequence.
ACCESSION BZ731737
VERSION BZ731737.1 GI:28705867
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 829)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGFAS29TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers

FEATURES
source
Query Match 4.6%; Score 23; DB 29; Length 820;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGCC 320
Db 506 CGCGAGTCTCGCGCGCGTGCC 484

RESULT 6
BZ731737/c
LOCUS BZ731737/c 829 bp DNA linear GSS 03-MAR-2003
DEFINITION OGFAS22TM ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0241P09,
genomic survey sequence.
ACCESSION BZ731737
VERSION BZ731737.1 GI:28705867
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 829)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGFAS29TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers

FEATURES
source
Query Match 4.6%; Score 23; DB 29; Length 742;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 CGCGAGTCTCGCGCGCGTGCC 320
Db 147 CGCGAGTCTCGCGCGCGTGCC 169

RESULT 5
CG297728/c
LOCUS CG297728/c 820 bp DNA linear GSS 25-AUG-2003
DEFINITION OG3BA37TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0766H01,
genomic survey sequence.

```

```

source
1. .829
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMBMa0241F09"
/clone_lib="ZM 0.7_1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 829;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCCTCGCGCGCGTCC 320
Db 415 CGCGAGTCCTCGCGCGCGTCC 393

RESULT 7
CC381148 857 bp DNA linear GSS 19-MAY-2003
LOCUS PHNW75TD ZM_0.6_1.0_KB Zea mays genomic clone ZMBMa494W05,
DEFINITION genomic survey sequence.
ACCESSION CC381148
VERSION CC381148.1 GI:30858013
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 857)
AUTHORS Whitelaw C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUHNW75TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .857
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMBMa0241F09"
/clone_lib="ZM 0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 857;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCCTCGCGCGCGTCC 320
Db 495 CGCGAGTCCTCGCGCGCGTCC 517

RESULT 8
BZ731730 858 bp DNA linear GSS 03-MAR-2003
LOCUS BZ731730
DEFINITION OGFAS29TC ZM_0.7_1.5_KB Zea mays genomic clone ZMBMa0241F09,

```

```

genomic survey sequence.
BZ731730
VERSION BZ731730.1 GI:28705853
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 858)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGFAS29TM
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .858
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMBMa0241F09"
/clone_lib="ZM 0.7_1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 4.6%; Score 23; DB 28; Length 858;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCCTCGCGCGCGTCC 320
Db 777 CGCGAGTCCTCGCGCGCGTCC 799

RESULT 9
CG092377 872 bp DNA linear GSS 20-AUG-2003
LOCUS PUFWT10TD ZM_0.6_1.0_KB Zea mays genomic clone ZMBMa0750B20,
DEFINITION genomic survey sequence.
ACCESSION CG092377
VERSION CG092377.1 GI:33974671
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 872)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUFWT10TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

```

```

FEATURES
  source      Location/Qualifiers
  1..872
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMBTA0750B20"
    /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
    Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 29; Length 872;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGGTGCC 320
Db 278 CGCGAGTCTCTCGCGCGGTGCC 300

RESULT 10
CG289214/c
LOCUS
DEFINITION
  CG289214 876 bp DNA linear GSS 25-AUG-2003
  OGMWKS6TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0617J15,
  genomic survey sequence.
ACCESSION
  CG289214
VERSION
  CG289214.1 GI:34203428
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 876)
  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: OGMWKS6TV
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TR
  Class: sheared ends.
  Location/Qualifiers
    1..876
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0617J15"
      /notes="Vector: PBCSk-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 29; Length 876;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGGTGCC 320
Db 413 CGCGAGTCTCTCGCGCGGTGCC 391

RESULT 11
CC381145/c
LOCUS
DEFINITION
  CC381145 900 bp DNA linear GSS 19-MAY-2003
  OGMWKS6TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0617J15,
  genomic survey sequence.
ACCESSION
  CC381145
VERSION
  CC381145.1 GI:30858008
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 900)
  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
  Bennetzen, J.
  Maize Genomics Consortium
  Unpublished (2003)
  Other_GSSs: PUHNW75TD
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TR
  Class: sheared ends.
  Location/Qualifiers
    1..900
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBTA494M05"
      /clone_lib="ZM 0.6 1.0 KB"
      /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 28; Length 900;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGGTGCC 320
Db 408 CGCGAGTCTCTCGCGCGGTGCC 386

RESULT 12
CG384209/c
LOCUS
DEFINITION
  CG384209 918 bp DNA linear GSS 26-AUG-2003
  OG0BR89TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0681P10,
  genomic survey sequence.
ACCESSION
  CG384209
VERSION
  CG384209.1 GI:34301476
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 918)
  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: OG0BR89TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
  Location/Qualifiers
    1..918
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0681P10"
      /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 28; Length 918;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGGTGCC 320
Db 408 CGCGAGTCTCTCGCGCGGTGCC 386

RESULT 13
CG384209/c
LOCUS
DEFINITION
  CG384209 918 bp DNA linear GSS 26-AUG-2003
  OG0BR89TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0681P10,
  genomic survey sequence.
ACCESSION
  CG384209
VERSION
  CG384209.1 GI:34301476
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 918)
  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other_GSSs: OG0BR89TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
  Location/Qualifiers
    1..918
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0681P10"
      /notes="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
Query Match      4.6%; Score 23; DB 28; Length 918;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 CGCGAGTCTCTCGCGCGGTGCC 320
Db 408 CGCGAGTCTCTCGCGCGGTGCC 386

```

```

FEATURES
  source
    Class: sheared ends.
    Location/Qualifiers
      1..918
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMWBMA061P10"
        /clone_lib="ZM_0.7_1.5_KB"
        /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
        methylation filtered genomic DNA library"
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 49;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 298 CGCGAGTCTCGCGCGCGTGC 320
  Db 878 CGCGAGTCTCGCGCGCGTGC 856
RESULT 13
LOCUS CG289224
DEFINITION OGMWKS6TV ZM_0.7_1.5_KB Zea mays genomic clone ZMWBMA061P15,
genomic survey sequence.
ACCESSION CG289224
VERSION CG289224.1 GI:34203438
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 974)
  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
  Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
  Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other GSSs: OGMWKS6TV
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
  Location/Qualifiers
    1..974
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMWBMA061P15"
      /clone_lib="ZM_0.7_1.5_KB"
      /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 49;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 298 CGCGAGTCTCGCGCGCGTGC 320
  Db 942 CGCGAGTCTCGCGCGCGTGC 964
RESULT 14
LOCUS AY103974
DEFINITION Zea mays PCO123560 mRNA sequence.
ACCESSION AY103974
VERSION AY103974.1 GI:21207052
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 1889)
  Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
  Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
  Maize Mapping Project/DuPont Consensus Sequences for Design of
  Overgo Probes
  Unpublished (2002)
  2 (bases 1 to 1889)
  Coe, E.H.
  Direct Submission
  Submitted (25-APR-2002) Maize Mapping Project, University of
  Missouri, Columbia, MO 65211, USA
  If you are interested in getting corresponding physical clones,
  these are publicly available from ZmDB and may be found by BLAST
  searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
  www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
  maize cDNA sequences is either Virginia Walbot, Stanford or Pat
  Schnable, Iowa State, then clones may be requested from ZmDB:
  www.zmdb.iastate.edu.
  Location/Qualifiers
    1..1889
      /organism="Zea mays"
      /mol_type="mRNA"
      /db_xref="MaizeDB:637184"
      /db_xref="taxon:4577"
      /clone_lib="Maize Mapping Project/DuPont Consensus
      Library"
      /note="this sequence is part of a project of EST
      assemblies resulting from the application of public
      configs to seed buPont contigs; this resource was
      assembled by DuPont as part of a collaboration for the
      overgo addressing of BACs in conjunction with the Maize
      Mapping Project"
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 52;
  Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 298 CGCGAGTCTCGCGCGCGTGC 320
  Db 565 CGCGAGTCTCGCGCGCGTGC 543
RESULT 15
LOCUS D15897
DEFINITION R1C1483A Rice callus Oryza sativa (japonica cultivar-group) cDNA
clone C1483A, mRNA sequence.
ACCESSION D15897
VERSION D15897.1 GI:287092
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
  1 (bases 1 to 340)
  Sasaki, T., Song, J., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H.,
  Nagasaki, H., Hori, M., Miya, M. and Murayama-Kayano, E.
  Toward cataloging all rice genes: large-scale sequencing of
  randomly chosen rice cDNAs from a callus cDNA library
  Plant J. 6 (4), 615-624 (1994)
  95078950
  MEDLINE
  
```



see Close TJ, Wing R, Kleinjohs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html) "

## ORIGIN

Query Match 4.2%; Score 21; DB 10; Length 1189;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 CTGGCCACAGCGCGCGGAA 87  
|||||  
Db 933 CTGGCCACAGCGCGCGGAA 913

## RESULT 18

CF770975  
LOCUS  
DEFINITION  
DSBF1\_12\_C09.g1\_A010 Drought-stressed before flowering Sorghum EST 20-OCT-2003  
bicolor cDNA clone DSBF1\_12\_C09\_A010 3', mRNA sequence.

ACCESSION  
CF770975 1 GI:37755892

VERSION  
EST.

KEYWORDS  
Sorghum bicolor (sorghum)

SOURCE  
Sorghum bicolor

ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE  
1 (bases 1 to 354)

AUTHORS  
Cordonnier-Pratt, M.-M., Zhang, D., McCartor, K., Nguyen, H.T. and Pratt, L.H.

TITLE  
An EST Database from Sorghum: Subtracted pre-flowering drought

stressed leaf tissues

UNPUBLISHED (2003)

CONTACT: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed at Texas Tech University by Deshui Zhang in

the laboratory of Dr. Henry Nguyen. Sequencing was done in the

Laboratory for Genomics and Bioinformatics, University of Georgia.

Sequence ends have been trimmed to exclude vector and regions below

phred quality 16. Three-prime sequences are presented as their

reverse complement and have been trimmed to exclude polyA.

Seq primer: T7 (TAATACGACTCACTATAGG)

POLYA=Yes.

FEATURES

Location/Qualifiers

1..354

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="Tx7000"

/db\_xref="taxon:4558"

/clone="DSBF1\_12\_C09\_A010"

/dev\_stage="Pre-flowering"

/lab\_host="ElectroMax DH10B (BRL)"

/clone\_lib="drought-stressed before flowering"

/notes="Organ: Leaf; Vector: pBluescriptSK-; Site: 1: XhoI;

Site 2: EcoRI; The library was prepared from polyA+ RNA

from leaves harvested from pre-flowering, drought-stressed

Sorghum bicolor, cv. TX7000. Double-stranded cDNA was

cloned unidirectionally using the Unizap system from

Stratagene. After amplification, the library was

subtracted by re-association hybridization. Inserts can be

excised with XhoI and EcoRI."

## ORIGIN

Query Match 4.0%; Score 20; DB 14; Length 354;  
Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 TCGGCGCCGTCGCGTCGCC 327  
|||||  
Db 19 TCGGCGCCGTCGCGTCGCC 38

## RESULT 19

CF759576

LOCUS

DEFINITION

bicolor cDNA clone DSAF1\_46\_H11\_A011 5', mRNA sequence.

ACCESSION

CF759576

VERSION

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor (sorghum)

Sorghum bicolor

REFERENCE

1 (bases 1 to 375)

AUTHORS

Cordonnier-Pratt, M.-M., Zhang, D., McCartor, K., Nguyen, H.T. and Pratt, L.H.

TITLE

An EST Database from Sorghum: Subtracted post-flowering drought

stressed leaf tissues

UNPUBLISHED (2003)

CONTACT: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed at Texas Tech University by Deshui Zhang and

Jianhang Jia in the laboratory of Dr. Henry Nguyen. Sequencing was

done in the Laboratory for Genomics and Bioinformatics, University

of Georgia. Sequence ends have been trimmed to exclude vector and

regions below phred quality 16. Three-prime sequences are presented

as their reverse complement and have been trimmed to exclude polyA.

Seq primer: JENREV (CAGGAACAGCTATGACC).

FEATURES

Location/Qualifiers

1..375

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="B35"

/db\_xref="taxon:4558"

/clone="DSAF1\_46\_H11\_A011"

/dev\_stage="Post-flowering"

/lab\_host="ElectroMax DH10B (BRL)"

/clone\_lib="Drought-stressed after flowering"

/notes="Organ: Leaf; Vector: pBluescriptSK-; Site: 1: XhoI;

Site 2: EcoRI; The library was prepared from polyA+ RNA

from leaves harvested from post-flowering,

drought-stressed Sorghum bicolor, cv. B35. Double-stranded

cDNA was cloned unidirectionally using the Unizap system

from Stratagene. After amplification, the library was

subtracted by re-association hybridization. Inserts can be

excised with XhoI and EcoRI."

## ORIGIN

Query Match 4.0%; Score 20; DB 14; Length 375;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 TCGGCGCCGTCGCGTCGCC 327  
|||||

Db 19 TCGGCGCCGTCGCGTCGCC 38

## RESULT 20

BM371243/c

LOCUS

DEFINITION

EBro04\_SQ003\_P12\_R root, 3 week, salt-stressed, cv Optic, EBro04

Hordeum vulgare subsp. vulgare cDNA clone EBro04\_SQ003\_P12 5', mRNA

sequence.

BM371243 491 bp mRNA linear EST 23-JUL-2002  
EBro04\_SQ003\_P12\_R root, 3 week, salt-stressed, cv Optic, EBro04  
Hordeum vulgare subsp. vulgare cDNA clone EBro04\_SQ003\_P12 5', mRNA  
sequence.



TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

FEATURES Location/Qualifiers  
source  
1..702  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0343M24"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 28; Length 702;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 138 TGCTCGGCTCGCGACACC 157  
|||||

## RESULT 23

BG962173/c 717 bp mRNA linear EST 12-JUN-2001  
LOCUS 602826940F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4981881 5',  
DEFINITION mRNA sequence.

ACCESSION BG962173  
VERSION BG962173.1 GI:14349810  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 717)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM10983 row: p column: 10  
High quality sequence stop: 249.

## FEATURES

source  
1..717  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4981881"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP Co24"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 4.0%; Score 20; DB 12; Length 717;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CGCGGCCACCGGCTCGGC 184  
|||||  
Db 98 CGCGGCCACCGGCTCGGC 79  
|||||

## RESULT 24

CC670138 770 bp DNA linear GSS 19-JUN-2003  
LOCUS OGLCF27TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0321E06,  
DEFINITION genomic survey sequence.

ACCESSION CC670138  
VERSION CC670138.1 GI:32074380  
KEYWORDS GSS.  
SOURCE Zea mays

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 770)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## TITLE

Consortium for Maize Genomics

## JOURNAL

## COMMENT

Unpublished (2002)

Other GSSs: OGLCF27TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

## FEATURES

source  
1..770  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0321E06"  
/clone\_lib="ZM 0.7 1.5 KB"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 770;  
Best Local Similarity 100.0%; Pred. No. 8.5e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 TGCTCGGCTCGCGACACC 378  
|||||  
Db 254 TGCTCGGCTCGCGACACC 273  
|||||

## RESULT 25

CG382731/c 783 bp DNA linear GSS 26-AUG-2003  
LOCUS OGLBK25TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0724E01,  
DEFINITION genomic survey sequence.

ACCESSION CG382731  
VERSION CG382731.1 GI:34299998  
KEYWORDS GSS.  
SOURCE Zea mays

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 783)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,



Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

# TITLE JOURNAL

COMMENT  
Unpublished (2002)

Other\_GSSs: OG1BK25TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

## FEATURES

source

Location/Qualifiers

1..783

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMBMa0724E01"

/clone\_lib="ZM\_0.7\_1.5\_KB"

/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 783;  
Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

QY 359 TGCTCGGCTCGCGACACC 378

|||||

Db 431 TGCTCGGCTCGCGACACC 412

## RESULT 26

CG382740

LOCUS

CG1BK25TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMBMa0724E01, linear GSS 26-AUG-2003

DEFINITION

genomic survey sequence.

ACCESSION

CG382740

VERSION

CG382740.1 GI:34300007

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 802)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,

Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OG1BK25TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..802

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMBMa0724E01"

/clone\_lib="ZM\_0.7\_1.5\_KB"

/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.0%; Score 20; DB 29; Length 802;

Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0;

QY 359 TGCTCGGCTCGCGACACC 378

|||||

Db 713 TGCTCGGCTCGCGACACC 732

## RESULT 27

BZ649718/c

LOCUS

OGCBQ11TC ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMBMa0135B21, linear GSS 29-JAN-2003

DEFINITION

genomic survey sequence.

ACCESSION

BZ649718

VERSION

BZ649718.1 GI:28114802

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 849)

Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,

Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OGCBQ11TM

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..849

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMBMa0135B21"

/clone\_lib="ZM\_0.7\_1.5\_KB"

/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 4.0%; Score 20; DB 28; Length 849;

Best Local Similarity 100.0%; Pred.No. 8.5e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0;

QY 410 CCGAAGCGGCGCGCGCC 429

|||||

Db 740 CCGAAGCGGCGCGCGCC 721

## RESULT 28

BE791387

LOCUS

BE791387F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3938540 5', linear EST 20-SEP-2000

DEFINITION

mRNA sequence.

ACCESSION

BE791387

VERSION

BE791387.1 GI:10212585

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1035)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LICM795 row: 0 column: 21  
High quality sequence stop: 444.

## FEATURES

source

Location/Qualifiers  
1. .1035  
/organism="Homo sapiens"  
/mol type="mRNA"  
/db xref="taxon:9606"  
/clone="IMAGE:3938540"  
/tissue type="small cell carcinoma"  
/cell line="MGC3"  
/lab host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 4.0%; Score 20; DB 10; Length 1035;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GGACGGCGGAGACGACC 295

|||||  
Db 776 GGACGGCGGAGACGACC 795

## RESULT 29

CB927303/c

LOCUS

DEFINITION ABAL\_14\_B04.g1 A012 Absciscic acid-treated seedlings Sorghum bicolor  
cDNA clone ABAL\_14\_B04\_A012 5', mRNA sequence.

ACCESSION

CB927303

VERSION

CB927303.1

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum bicolor (sorghum)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

Clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 112)

Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Buchanan,C.D., Eastman,A. and Pratt,L.H.

An EST database from Sorghum: ABAL-treated seedlings

Unpublished (2003)

Other ESTs: ABAL\_14\_B04.b1\_A012

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A &amp; M University;

sequencing done in the Laboratory for Genomics and Bioinformatics;

University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGAGCTCG).

## FEATURES

source

Location/Qualifiers  
1. .112  
/organism="Sorghum bicolor"  
/mol type="mRNA"  
/cultivar="IS3620C"  
/db xref="taxon:4558"  
/clone="ABAL\_14\_B04\_A012"  
/lab host="DH10B-T1 phage-resistant E. coli"  
/clone lib="Absciscic acid-treated seedlings"  
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. After 12 days, medium was supplemented with 1 mM absciscic acid (ABA), while leaves were misted with a solution of 1 mM ABA. Roots and leaves were harvested after 3, 6, 12, and 24 hr and material from all time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGATG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

## ORIGIN

Query Match 3.8%; Score 19; DB 14; Length 112;

Best Local Similarity 100.0%; Pred. No. 1.8e+03;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 CGGCCCGCGCGCGACGA 434

|||||

Db 89 CGGCCCGCGCGCGACGA 71

## RESULT 30

BF822655

LOCUS

DEFINITION

CM2-RT0009-091200-605-b05 RT0009 Homo sapiens

ACCESSION

BF822655

VERSION

BF822655.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Chordata; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 211)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

MEDLINE

10737800

PUBMED

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&amp;tl2=CM2-RT0009-

091200-605-b05&amp;tl3=2000-12-09&amp;tl4=1)

Seq primer: puc 18 forward

High quality sequence start: 24

High quality sequence stop: 158.

```

FEATURES
  source
    Location/Qualifiers
      1..211
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="RT0009"
        /note="Organ: kidney tumor; Vector: puc18; Site 1: SmaI;
        Site 2: SmaI; A mini-library was made by cloning products
        derived from ORESTES PCR (U.S. Letters Patent application
        No. 196,716 - Ludwig Institute for Cancer Research)
        profiles into the pUC 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
      3.8%; Score 19; DB 10; Length 211;
      Best Local Similarity 100.0%; Pred. No. 1.9e+03;
      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 27 CATCTGCTGAAGCTGGTC 45
      |||||
  Db 134 CATCTGCTGAAGCTGGTC 152

RESULT 31
LOCUS AI352333
DEFINITION qt19g03.x1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1948084 3',
similar to SW:UROC_RAT P55090 UROCORTIN PRECURSOR ;, mRNA sequence.
ACCESSION AI352333
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbtp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers
  1..219
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:1948084"
    /tissue_type="pooled germ cell tumors"
    /lab_host="DH10B"
    /clone_lib="NCI CGAP GC4"
    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; 1st strand cDNA was prepared from 3 pooled
    germ cell tumors, and was then primed with a Not I -
    oligo(dT) primer. Double-stranded cDNA was ligated to Eco
    RI adaptors (Pharmacia), digested with Not I and cloned
    into the Not I and Eco RI sites of the modified pT7T3
    vector. Library is normalized. Library was constructed by
    Bento Soares and M. Fatima Bonaldo."

FEATURES
  source
    Location/Qualifiers
      1..219
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:1948084"
        /tissue_type="pooled germ cell tumors"
        /lab_host="DH10B"
        /clone_lib="NCI CGAP GC4"
        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
        polylinker; 1st strand cDNA was prepared from 3 pooled
        germ cell tumors, and was then primed with a Not I -
        oligo(dT) primer. Double-stranded cDNA was ligated to Eco
        RI adaptors (Pharmacia), digested with Not I and cloned
        into the Not I and Eco RI sites of the modified pT7T3
        vector. Library is normalized. Library was constructed by
        Bento Soares and M. Fatima Bonaldo."

  Query Match 3.8%; Score 19; DB 9; Length 219;
  Best Local Similarity 100.0%; Pred. No. 1.9e+03;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 12 CGCGGAGCAGAACCGCATC 30
      |||||
  Db 72 CGCGGAGCAGAACCGCATC 90

RESULT 33
LOCUS BF773138
DEFINITION CM4-IT0042-151200-605-b05 IT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF773138
VERSION BF773138.1 GI:12121038

  Query Match 3.8%; Score 19; DB 9; Length 219;
  Best Local Similarity 100.0%; Pred. No. 1.9e+03;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 12 CGCGGAGCAGAACCGCATC 30
      |||||
  Db 72 CGCGGAGCAGAACCGCATC 90

RESULT 33
LOCUS BF773138
DEFINITION CM4-IT0042-151200-605-b05 IT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF773138
VERSION BF773138.1 GI:12121038

```

```

FEATURES
  source
    Location/Qualifiers
      1..219
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:2238826"
        /tissue_type="pooled germ cell tumors"
        /lab_host="DH10B"
        /clone_lib="NCI CGAP GC6"
        /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
        polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
        from the normalized library NCI CGAP GC4 was prepared, and
        ss circles were made in vitro. Following HAP purification,
        this DNA was used as tracer in a subtractive hybridization
        reaction. The driver was PCR-amplified cDNAs from a pool
        of 5,000 clones made from the same library (cloneIDs
        1257096-1258631, 1469064-1470983, and 1475592-1476743).
        Subtraction by Bento Soares and M. Fatima Bonaldo."

  Query Match 3.8%; Score 19; DB 9; Length 219;
  Best Local Similarity 100.0%; Pred. No. 1.9e+03;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 12 CGCGGAGCAGAACCGCATC 30
      |||||
  Db 72 CGCGGAGCAGAACCGCATC 90

RESULT 33
LOCUS BF773138
DEFINITION CM4-IT0042-151200-605-b05 IT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF773138
VERSION BF773138.1 GI:12121038

```

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 220)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-IT0042-
151200-605-b05&t3=2000-12-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 32
High quality sequence stop: 166.
FEATURES
source
1..220
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="IT0042"
/note="Organ: epid tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORFESTS PCR (O.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
ORIGIN
Query Match 3.8%; Score 19; DB 10; Length 220;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCCTGCTGAAGCTGGTC 45
Db 142 CATCCTGCTGAAGCTGGTC 160

REFERENCE 34
AI971801
LOCUS w28dl1.x1 NCI CGAP Ov18 Homo sapiens cDNA clone IMAGE:2530869 3'
DEFINITION similar to TR:O77577 O77577 UROCORTIN PRECURSOR 1, mRNA sequence.
ACCESSION AI971801
VERSION AI971801.1 GI:5768627
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 251)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: CGAP@remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone description: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 239.
FEATURES
source
1..251
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2530869"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI-CGAP_Ov18"
/note="Organ: ovary; Vector: pT73D-Pac (pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGCGCGCGACATTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match 3.8%; Score 19; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCGGAGCAGACCGCATC 30
Db 72 CGCGGAGCAGACCGCATC 90

RESULT 35
AI716696/c
LOCUS AI716696 346 bp mRNA linear EST 10-JUN-1999
DEFINITION UI-R-Y0-acf-b-01-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
UI-R-Y0-acf-b-01-0-UI 3', mRNA sequence.
ACCESSION AI716696
VERSION AI716696.1 GI:5033952
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 346)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 8889548
PUBMED
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)

```

Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

Location/Qualifiers  
1. .346  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-Y0-acf-b-01-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-Y0"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-Y0  
library is a subtracted library derived from an  
individually-tagged normalized whole-eye (minus the lens)  
library. The driver for the subtraction consisted of a  
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,  
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of  
3-5 nucleotides present between the Not I site and the  
oligo-dT track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-Y0) was constructed as follows: PCR  
amplified cDNA inserts from previous library clones from  
which 3' ESTs had been derived were used as a driver in a  
hybridization with the normalized whole-eye library in  
the form of single-stranded circles. The remaining  
single-stranded circles (subtracted library) was purified  
by hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-Y0  
library. This procedure has been previously described  
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,  
1996)

TAG TISSUE=Eye  
TAG\_LIB=UI-R-Y0  
TAG\_SEQ=CAATG

## ORIGIN

Query Match 3.8%; Score 19; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0;  
QY 101 GCCAGCGTTCAGGAGGT 119  
|||||  
DB 149 GCCAGCGTTCAGGAGGT 131

## RESULT 36

BE415413

LOCUS

DEFINITION

BE415413 349 bp mRNA linear EST 24-JUL-2000  
clone MWL030.B02000309 ITEC MWL Wheat Root Library Triticum aestivum CDNA

ACCESSION

BE415413

VERSION

BE415413.1

KEYWORDS

EST.

SOURCE

Triticum aestivum (bread wheat)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

REFERENCE

AUTHORS

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,  
Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,  
Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,  
Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,  
Ogihara, Y., Pecchioni, N., Qaluset, C., Schuch, W., Selvaraj, G.,  
Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.  
International Triticaceae EST Cooperative (ITEC): Production of  
Expressed Sequence Tags for Species of the Triticeae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT

TITLE

JOURNAL

COMMENT

Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cnet.com  
International Triticaceae EST Cooperative (ITEC)  
http://wheat.pw.usda.gov/genome.

## FEATURES

source

Location/Qualifiers  
1. .349  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Atlas"  
/db\_xref="taxon:4565"  
/clone="MWL030.B02"  
/tissue\_type="root"  
/dev\_stage="8 day old"  
/clone\_lib="ITEC MWL Wheat Root Library"  
/note="Vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average  
insert size."

## ORIGIN

Query Match 3.8%; Score 19; DB 10; Length 349;  
Best Local Similarity 100.0%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0;  
QY 31 CTGCTGAAGCTGCTCGCG 49  
|||||  
DB 83 CTGCTGAAGCTGCTCGCG 101

## RESULT 37

BF941825

LOCUS

DEFINITION

BF941825 356 bp mRNA linear EST 30-MAR-2001  
nad93h05.xl NCI CGAP Ovl8 Homo sapiens cDNA clone IMAGE:3433569.3  
similar to SW:UROC\_MOUSE P81615 UROCORTIN PRECURSOR. [1]; mRNA

ACCESSION

BF941825

VERSION

BF941825.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

JOURNAL

COMMENT

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Benito  
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 334.  
Location/Qualifiers  
1. .356  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3433569"  
/tissue\_type="fibrotheoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NCI CGAP\_Ovl8"  
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTCGAGCGCGCGACATTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Eco RI adaptors

## FEATURES

source

Location/Qualifiers  
1. .356  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3433569"  
/tissue\_type="fibrotheoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NCI CGAP\_Ovl8"  
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTCGAGCGCGCGACATTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

```

ORIGIN
Query Match          3.8%; Score 19; DB 10; Length 356;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CGCCGAGCAGACCGCATC 30
      |||||
Db 72 CGCCGAGCAGACCGCATC 90

RESULT 38
CA646451
LOCUS
DEFINITION          358 bp mRNA linear EST 23-NOV-2002
wreln.pk0095.h11 wreln Triticum aestivum cDNA clone
ACCESSION          CA646451
VERSION            CA646451.1 GI:25224747
KEYWORDS
SOURCE             Triticum aestivum (bread wheat)
ORGANISM           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
REFERENCE
AUTHORS            Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE              DuPont Wheat cDNA Sequence
JOURNAL            Unpublished (2002)
COMMENT            Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES
source
1..358
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wreln.pk0095.h11"
/tissue_type="root"
/clone_lib="wreln"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site 2:
XhoI; wheat (Triticum aestivum L.) root; normalized from
wreln library"

ORIGIN
Query Match          3.8%; Score 19; DB 14; Length 358;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 CTGCTGAAGCTGTCGCG 49
      |||||
Db 135 CTGCTGAAGCTGTCGCG 153

RESULT 39
H24268
LOCUS
DEFINITION          379 bp mRNA linear EST 06-JUL-1995
vm55c11.r1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:52283 5' similar to gb:X63692 DNA (HUMAN); mRNA sequence.
ACCESSION          H24268
VERSION            H24268.1 GI:892963
KEYWORDS            EST.
SOURCE             Homo sapiens (human)
ORGANISM           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
TITLE              The WashU-Merck EST Project
JOURNAL            Unpublished (1995)
COMMENT            Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1983
High quality sequence stops: 370 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1983 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 370.
Location/Qualifiers
1..379
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:425319"
/db_xref="taxon:9606"
/clone="IMAGE:52383"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain INIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5,
AAGTGAAGATTCGCGCGAGGATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match          3.8%; Score 19; DB 14; Length 379;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 CATCTGCTGAAGCTGGTC 45
      |||||
Db 297 CATCTGCTGAAGCTGGTC 315

RESULT 40
BF523813
LOCUS
DEFINITION          384 bp mRNA linear EST 11-DEC-2000
UI-R-Y0-acf-b-01-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone
ACCESSION          BF523813
VERSION            BF523813.1 GI:11631780
KEYWORDS            EST.
SOURCE             Rattus norvegicus (Norway rat)
ORGANISM           Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS            1 (bases 1 to 384)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE              Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL            Genome Res. 6 (9), 791-806 (1996)
MEDLINE            97044477

```

**PUBMED COMMENT**  
 8899548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
 clones will be available through Research Genetics (www.resgen.com)  
 This clone is also available through the I.M.A.G.E. Consortium at  
 LLNL (info@image.llnl.gov). IMAGE ID= 1791476  
 Seq primer: M13 Forward.

**FEATURES**  
 source  
 1..384  
 /location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-Y0-act-b-01-0-UI"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="UI-R-Y0"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-Y0  
 library is a subtracted library derived from an  
 individually-tagged normalized whole-eye (minus the lens)  
 library. The driver for the subtraction consisted of a  
 pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,  
 UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of  
 3-5 nucleotides present between the Not I site and the  
 oligo-dT track which allows identification of the library  
 of origin of a clone within the mixture. The subtracted  
 library (UI-R-Y0) was constructed as follows: PCR  
 amplified cDNA inserts from previous library clones from  
 which 3' ESTs had been derived were used as a driver in a  
 hybridization with the normalized whole-eye library in  
 the form of single-stranded circles. The remaining  
 single-stranded circles (subtracted library) was purified  
 by hydroxyapatite column chromatography, converted to  
 double-stranded circles and electroporated into DH10B  
 bacteria (Life Technologies) to generate the UI-R-Y0  
 library. This procedure has been previously described  
 (Bonaldo, Lemmon and Soares, Genome Research 6: 791-806,  
 1996)"

**ORIGIN**  
 Query Match 3.8%; Score 19; DB 10; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Qy** 101 GCCAGGCGTTCAGGAGGT 119  
 |||||  
 |||||

**Db** 212 GCCAGGCGTTCAGGAGGT 230  
 |||||  
 |||||

**RESULT 41**  
**CA614966**  
**LOCUS** 399 bp mRNA linear EST 23-NOV-2002  
**DEFINITION** wrl.pk149.h5 wrl Triticum aestivum cDNA clone wrl.pk149.h5 5' end,  
 mRNA sequence.  
**ACCESSION** CA614966  
**VERSION** CA614966.1 GI:25193277  
**KEYWORDS** EST.  
**SOURCE** Triticum aestivum (bread wheat)  
**ORGANISM** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (bases 1 to 399)  
**REFERENCE** Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,  
**AUTHORS** Miao,G., Caraher,N. and Hanafey,M.K.  
**TITLE** DuPont Wheat cDNA Sequence

**JOURNAL COMMENT**  
 Unpublished (2002)  
 Contact: Scott V. Tingey  
 Crop Genetics  
 B. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.

**FEATURES**  
 source  
 1..399  
 /location/Qualifiers  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone="wrl.pk149.h5"  
 /tissue\_type="root"  
 /clone\_lib="wrl"  
 /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:  
 XhoI; Wheat (Triticum aestivum L.) root; 7 day old  
 seedling, light grown"

**ORIGIN**  
 Query Match 3.8%; Score 19; DB 14; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Qy** 31 CTGCTGAAGCTGTCGCGC 49  
 |||||  
 |||||

**Db** 14 CTGCTGAAGCTGTCGCGC 32  
 |||||  
 |||||

**RESULT 42**  
**BE415414**  
**LOCUS** 410 bp mRNA linear EST 24-JUL-2000  
**DEFINITION** MWL030.B03000309 ITEC MWL Wheat Root Library Triticum aestivum cDNA  
 clone MWL030.B03, mRNA sequence.  
**ACCESSION** BE415414  
**VERSION** BE415414.1 GI:9413260  
**KEYWORDS** EST.  
**SOURCE** Triticum aestivum (bread wheat)  
**ORGANISM** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (bases 1 to 410)  
**REFERENCE** Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T.,  
**AUTHORS** Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A.,  
 Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J.,  
 Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P.,  
 Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G.,  
 Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.  
 International Triticeae EST Cooperative (ITEC): Production of  
 Expressed Sequence Tags for Species of the Triticeae  
 Unpublished (2000)  
 Contact: Warburton M  
 Applied Biotechnology Center, CIMMYT  
 Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
 Tel: 52-5-7269091 ext 1381  
 Fax: 52-5-7267558/59  
 Email: mwarburton@cnet.com  
 International Triticeae EST Cooperative (ITEC)  
 http://wheat.pw.usda.gov/genome.

**TITLE** Location/Qualifiers  
 1..410  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Atlas"  
 /db\_xref="taxon:4565"  
 /clone="MWL030.B03"  
 /tissue\_type="root"  
 /dev\_stage="8 day old"  
 /clone\_lib="ITEC MWL Wheat Root Library"  
 /note="Vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average